EXPRESSED SEQUENCES OF *ARABIDOPSIS THALIANA*

CROSS –REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application 60/178,512 Filed January 27, 2000.

FIELD OF INVENTION

The invention is in the field of polynucleotide sequences of a plant, particularly sequences expressed in arabidopsis thaliana.

BACKGROUND OF THE INVENTION

Plants and plant products have vast commercial importance in a wide variety of areas including food crops for human and animal consumption, flavor enhancers for food, and production of specialty chemicals for use in products such as medicaments and fragrances. In considering food crops for humans and livestock, genes such as those involved in a plant's resistance to insects, plant viruses, and fungi; genes involved in pollination; and genes whose products enhance the nutritional value of the food, are of major importance. A number of such genes have been described, see, for example, McCaskill and Croteau (1999) Nature Biotechnol. 17:31-36.

Despite recent advances in methods for identification, cloning, and characterization of genes, much remains to be learned about plant physiology in general, including how plants produce many of the above-mentioned products; mechanisms for resistance to herbicides, insects, plant viruses, fungi; elucidation of genes involved in specific biosynthetic pathways; and genes involved in environmental tolerance, *e.g.*, salt tolerance, drought tolerance, or tolerance to anaerobic conditions.

Arabidopsis thaliana is a model system for genetic, molecular and biochemical studies of higher plants. Features of this plant that make it a model system for genetic and molecular biology research include a small genome size, organized into

five chromosomes and containing an estimated 20,000 genes, a rapid life cycle, prolific seed production and, since it is small, it can easily be cultivation in limited space. *A. thaliana* is a member of the mustard family (*Brassicaceae*) with a broad natural distribution throughout Europe, Asia, and North America. Many different ecotypes have been collected from natural populations and are available for experimental analysis. The entire life cycle, including seed germination, formation of a rosette plant, bolting of the main stem, flowering, and maturation of the first seeds, is completed in 6 weeks. A large number of mutant lines are available that affect nearly all aspects of its growth. These features greatly facilitate the isolation of fundamentally interesting and potentially important genes for agronomic development

Most gene products from higher plants exhibit adequate sequence similarity to deduced amino acid sequences of other plant genes to permit assignment of probable gene function, if it is known, in any higher plant. It is likely that there will be very few protein-encoding angiosperm genes that do not have orthologs or paralogs in *Arabidopsis*. The developmental diversity of higher plants may be largely due to changes in the cis-regulatory sequences of transcriptional regulators and not in coding sequences.

Many advances reported over the past few years offer clear evidence that this plant is not only a very important model species for basic research, but also extremely valuable for applied plant scientists and plant breeders. Knowledge gained from *Arabidopsis* can be used directly to develop desired traits in plants of other species.

Relevant Literature

Cold Spring Harbor Monograph 27 (1994) E.M. Meyerowitz and C.R. Somerville, eds. (CSH Laboratory Press). Annual Plant Reviews, Vol. 1: Arabidopsis (1998) M. Anderson and J.A. Roberts, eds. (CRC Press). Methods in Molecular Biology: Arabidopsis Protocols, Vol. 82 (1997) J.M. Martinez-Zapater and J. Salinas, eds. (CRC Press).

Mayer *et al* (1999) Nature **402**(6763):769-77; "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana". Lin *et al.* (1999) **402**(6763):761-8,

"Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana". Meinke *et al.* (1998) Science **282**:662-682, "*Arabidopsis thaliana*: a model plant for genome analysis". Somerville and Somerville (1999) Science **285**:380-383, "Plant functional genomics". Mozo *et al.* (1999) Nat. Genet. **22**:271-275, "A complete BAC-based physical map of the *Arabidopsis thaliana* genome".

SUMMARY OF THE INVENTION

Novel nucleic acid sequences of *Arabidopsis thaliana,* their encoded polypeptides and variants thereof, genes corresponding to these nucleic acids, and proteins expressed by the genes, are provided.

The invention also provides diagnostic, prophylactic and therapeutic agents employing such novel nucleic acids, their corresponding genes or gene products, including expression constructs, probes, antisense constructs, and the like. The genetic sequences may also be used for the genetic manipulation of plant cells, particularly dicotyledonous plants. The encoded gene products and modified organisms are useful for introducing or improving disease resistance and stress tolerance into plants; screening of biologically active agents, *e.g.* fungicides, *etc.*; for elucidating biochemical pathways; and the like.

In one embodiment of the invention, a nucleic acid is provided that comprises a start codon; an optional intervening sequence; a coding sequence capable of hybridizing under stringent conditions as set forth in SEQ ID NO:1 to 999; and an optional terminal sequence, wherein at least one of said optional sequences is present. Such a nucleic acid may correspond to naturally occurring *Arabidopsis* expressed sequences.

DETAILED DESCRIPTION OF THE INVENTION

Novel nucleic acid sequences from *Arabidopsis thaliana*, their encoded polypeptides and variants thereof, genes corresponding to these nucleic acids and proteins expressed by the genes are provided. The invention also provides agents employing such novel nucleic acids, their corresponding genes or gene products,

including expression constructs, probes, antisense constructs, and the like. The nucleotide sequences are provided in the attached SEQLIST.

Sequences include, but are not limited to, sequences that encode resistance proteins; sequences that encode tolerance factors; sequences encoding proteins or other factors that are involved, directly or indirectly in biochemical pathways such as metabolic or biosynthetic pathways, sequences involved in signal transduction, sequences involved in the regulation of gene expression, structural genes, and the like. Biosynthetic pathways of interest include, but are not limited to, biosynthetic pathways whose product (which may be an end product or an intermediate) is of commercial, nutritional, or medicinal value.

The sequences may be used in screening assays of various plant strains to determine the strains that are best capable of withstanding a particular disease or environmental stress. Sequences encoding activators and resistance proteins may be introduced into plants that are deficient in these sequences. Alternatively, the sequences may be introduced under the control of promoters that are convenient for induction of expression. The protein products may be used in screening programs for insecticides, fungicides and antibiotics to determine agents that mimic or enhance the resistance proteins. Such agents may be used in improved methods of treating crops to prevent or treat disease. The protein products may also be used in screening programs to identify agents which mimic or enhance the action of tolerance factors. Such agents may be used in improved methods of treating crops to enhance their tolerance to environmental stresses.

Still other embodiments of the invention provide methods for enhancing or inhibiting production of a biosynthetic product in a plant by introducing a nucleic acid of the invention into a plant cell, where the nucleic acid comprises sequences encoding a factor which is involved, directly or indirectly in a biosynthetic pathway whose products are of commercial, nutritional, or medicinal value include any factor, usually a protein or peptide, which regulates such a biosynthetic pathway; which is an intermediate in such a biosynthetic pathway; or which in itself is a product that increases the nutritional value of a food product; or which is a medicinal product; or which is any product of commercial value.

Transgenic plants containing the antisense nucleic acids of the invention are useful for identifying other mediators that may induce expression of proteins of interest; for establishing the extent to which any specific insect and/or pathogen is responsible for damage of a particular plant; for identifying other mediators that may enhance or induce tolerance to environmental stress; for identifying factors involved in biosynthetic pathways of nutritional, commercial, or medicinal value; or for identifying products of nutritional, commercial, or medicinal value.

In still other embodiments, the invention provides transgenic plants constructed by introducing a subject nucleic acid of the invention into a plant cell, and growing the cell into a callus and then into a plant; or, alternatively by breeding a transgenic plant from the subject process with a second plant to form an F1 or higher hybrid. The subject transgenic plants and progeny are used as crops for their enhanced disease resistance, enhanced traits of interest, for example size or flavor of fruit, length of growth cycle, *etc.*, or for screening programs, *e.g.* to determine more effective insecticides, *etc*; used as crops which exhibit enhanced tolerance environmental stress; or used to produce a factor.

Those skilled in the art will recognize the agricultural advantages inherent in plants constructed to have either increased or decreased expression of resistance proteins; or increased or decreased tolerance to environmental factors; or which produce or over-produce one or more factors involved in a biosynthetic pathway whose product is of commercial, nutritional, or medicinal value. For example, such plants may have increased resistance to attack by predators, insects, pathogens, microorganisms, herbivores, mechanical damage and the like; may be more tolerant to environmental stress, *e.g.* may be better able to withstand drought conditions, freezing, and the like; or may produce a product not normally made in the plant, or may produce a product in higher than normal amounts, where the product has commercial, nutritional, or medicinal value. Plants which may be useful include dicotyledons and monocotyledons. Representative examples of plants in which the provided sequences may be useful include tomato, potato, tobacco, cotton, soybean, alfalfa, rape, and the like. Monocotyledons, more particularly grasses (*Poaceae* family) of interest, include, without limitation, *Avena sativa* (oat); *Avena strigosa*

(black oat); *Elymus* (wild rye); *Hordeum sp.* including *Hordeum vulgare* (barley); *Oryza sp.*, including *Oryza glaberrima* (African rice); *Oryza longistaminata* (long-staminate rice); *Pennisetum americanum* (pearl millet); *Sorghum sp.* (sorghum); *Triticum sp.*, including *Triticum aestivum* (common wheat); *Triticum durum* (durum wheat); *Zea mays* (corn); *etc.*

NUCLEIC ACID COMPOSITIONS

The following detailed description describes the nucleic acid compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these nucleic acids and genes; identification of structural motifs of the nucleic acids and genes; identification of the function of a gene product encoded by a gene corresponding to a nucleic acid of the invention; use of the provided nucleic acids as probes, in mapping, and in diagnosis; use of the corresponding polypeptides and other gene products to raise antibodies; use of the nucleic acids in genetic modification of plant and other species; and use of the nucleic acids, their encoded gene products, and modified organisms, for screening and diagnostic purposes.

The scope of the invention with respect to nucleic acid compositions includes, but is not necessarily limited to, nucleic acids having a sequence set forth in any one of SEQ ID NOS:1-999; nucleic acids that hybridize the provided sequences under stringent conditions; genes corresponding to the provided nucleic acids; variants of the provided nucleic acids and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product.

In one embodiment, the sequences of the invention provide a polypeptide coding sequence. The polypeptide coding sequence may correspond to a naturally expressed mRNA in Arabidopsis or other species, or may encode a fusion protein between one of the provided sequences and an exogenous protein coding sequence. The coding sequence is characterized by an ATG start codon, a lack of stop codons in-frame with the ATG, and a termination codon, that is, a continuous open frame is provided between the start and the stop codon. The sequence contained between the start and the stop codon will comprise a sequence capable of hybridizing under

stringent conditions to a sequence set for in SEQ ID NO:1-999, and may comprise the sequence set forth in the Seqlist.

Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here.

The invention features nucleic acids that are derived from *Arabidopsis thaliana*. Novel nucleic acid compositions of the invention of particular interest comprise a sequence set forth in any one of SEQ ID NOS:1-999 or an identifying sequence thereof. An "identifying sequence" is a contiguous sequence of residues at least about 10 nt to about 20 nt in length, usually at least about 50 nt to about 100 nt in length, that uniquely identifies a nucleic acid sequence, e.g., exhibits less than 90%, usually less than about 80% to about 85% sequence identity to any contiguous nucleotide sequence of more than about 20 nt. Thus, the subject novel nucleic acid compositions include full length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of SEQ ID NOS:1-999.

The nucleic acids of the invention also include nucleic acids having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M NaCl/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under stringent conditions, for example, at 50°C or higher and 0.1XSSC (9 mM NaCl/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see U.S. Patent No. 5,707,829. Nucleic acids that are substantially identical to the provided nucleic acid sequences, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided nucleic acid sequences (SEQ ID NOS:1-999) under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, particularly grasses as previously described.

Preferably, hybridization is performed using at least 15 contiguous nucleotides of at least one of SEQ ID NOS:1-999. The probe will preferentially hybridize with a nucleic acid or mRNA comprising the complementary sequence, allowing the

identification and retrieval of the nucleic acids of the biological material that uniquely hybridize to the selected probe. Probes of more than 15 nucleotides can be used, e.g. probes of from about 18 nucleotides up to the entire length of the provided nucleic acid sequences, but 15 nucleotides generally represents sufficient sequence for unique identification.

The nucleic acids of the invention also include naturally occurring variants of the nucleotide sequences, *e.g.* degenerate variants, allelic variants, *etc.* Variants of the nucleic acids of the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions For example, by using appropriate wash conditions, variants of the nucleic acids of the invention can be identified where the allelic variant exhibits at most about 25-30% base pair mismatches relative to the selected nucleic acid probe. In general, allelic variants contain 5-25% base pair mismatches, and can contain as little as even 2-5%, or 1-2% base pair mismatches, as well as a single base-pair mismatch.

The invention also encompasses homologs corresponding to the nucleic acids of SEQ ID NOS:1-999, where the source of homologous genes can be any related species, usually within the same genus or group. Homologs have substantial sequence similarity, *e.g.* at least 75% sequence identity, usually at least 90%, more usually at least 95% between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as BLAST, described in Altschul et al., J. Mol. Biol. (1990) 215:403-10.

In general, variants of the invention have a sequence identity greater than at least about 65%, preferably at least about 75%, more preferably at least about 85%, and can be greater than at least about 90% or more as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating

percent identity is the Smith-Waterman algorithm, using the following. Global DNA sequence identity must be greater than 65% as determined by the Smith-Wateman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty, 12; and gap extention penalty, 1.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein. The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide of the invention.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kb or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for expression.

The nucleic acid compositions of the subject invention can encode all or a part of the subject expressed polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, etc. Isolated nucleic acids and nucleic acid fragments of the invention comprise at least about 15 up to about 100 contiguous nucleotides, or up to the

complete sequence provided in SEQ ID NOS:1-999. For the most part, fragments will be of at least 15 nt, usually at least 18 nt or 25 nt, and up to at least about 50 contiguous nt in length or more.

Probes specific to the nucleic acids of the invention can be generated using the nucleic acid sequences disclosed in SEQ ID NOS:1-999 and the fragments as described above. The probes can be synthesized chemically or can be generated from longer nucleic acids using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of a nucleic acid of one of SEQ ID NOS:1-999. More preferably, probes are designed based on a contiguous sequence of one of the subject nucleic acids that remain unmasked following application of a masking program for masking low complexity (e.g., XBLAST) to the sequence., *i.e.* one would select an unmasked region, as indicated by the nucleic acids outside the poly-n stretches of the masked sequence produced by the masking program.

The nucleic acids of the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the nucleic acids, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", e.g., flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The nucleic acids of the invention can be provided as a linear molecule or within a circular molecule. They can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. They can be regulated by their own or by other regulatory sequences, as is known in the art. The nucleic acids of the invention can be introduced into suitable host cells using a variety of techniques which are available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The subject nucleic acid compositions can be used to, for example, produce polypeptides, as probes for the detection of mRNA of the invention in biological samples, *e.g.* extracts of cells, to generate additional copies of the nucleic acids, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of the nucleic acid sequences as shown in SEQ ID NOS:1-999 or variants thereof in a sample. These and other uses are described in more detail below.

USE OF NUCLEIC ACIDS AS CODING SEQUENCES

Naturally occurring Arabidopsis polypeptides or fragments thereof are encoded by the provided nucleic acids. Methods are known in the art to determine whether the complete native protein is encoded by a candidate nucleic acid sequence. Where the provided sequence encodes a fragment of a polypeptide, methods known in the art may be used to determine the remaining sequence. These approaches may utilize a bioinformatics approach, a cloning approach, extension of mRNA species, *etc.*

Substantial genomic sequence is available for Arabidopsis, and may be exploited for determining the complete coding sequence corresponding to the provided sequences. The region of the chromosome to which a given sequence is located may be determined by hybridization or by database searching. The genomic sequence is then searched upstream and downstream for the presence of intron/exon boundaries, and for motifs characteristic of transcriptional start and stop sequences, for example by using Genscan (Burge and Karlin (1997) J. Mol. Biol. **268**:78-94); or GRAIL (Uberbacher and Mural (1991) P.N.A.S. **88**:11261-1265).

Alternatively, nucleic acid having a sequence of one of SEQ ID NOS:1-999, or an identifying fragment thereof, is used as a hybridization probe to complementary molecules in a cDNA library using probe design methods, cloning methods, and clone selection techniques as known in the art. Libraries of cDNA are made from selected cells. The cells may be those of *A. thaliana*, or of related species. In some

cases it will be desirable to select cells from a particular stage, *e.g.* seeds, leaves, infected cells, *etc.*

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY; and Current Protocols in Molecular Biology, (1987 and updates) Ausubel et al., eds. The cDNA can be prepared by using primers based on sequence from SEQ ID NOS:1-999. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided nucleic acids, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. In order to obtain additional sequences 5' to the end of a partial cDNA, 5' RACE (PCR Protocols: A Guide to Methods and Applications, (1990) Academic Press, Inc.) may be performed.

Genomic DNA is isolated using the provided nucleic acids in a manner similar to the isolation of full-length cDNAs. Briefly, the provided nucleic acids, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the nucleic acids of the invention, but this is not essential. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook et al., 9.4-9.30. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook *et al.*, such that adjacent and overlapping fragments of genomic DNA are isolated. These are

mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

PCR methods may be used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant nucleic acids. Such PCR methods include gene trapping and RACE methods. Gene trapping entails inserting a member of a cDNA library into a vector. The vector then is denatured to produce single stranded molecules. Next, a substrate-bound probe, such a biotinylated oligo, is used to trap cDNA inserts of interest. Biotinylated probes can be linked to an avidin-bound solid substrate. PCR methods can be used to amplify the trapped cDNA. To trap sequences corresponding to the full length genes, the labeled probe sequence is based on the nucleic acid sequences of the invention. Random primers or primers specific to the library vector can be used to amplify the trapped cDNA. Such gene trapping techniques are described in Gruber *et al.*, WO 95/04745 and Gruber *et al.*, U.S. Pat. No. 5,500,356. Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA.

"Rapid amplification of cDNA ends", or RACE, is a PCR method of amplifying cDNAs from a number of different RNAs. The cDNAs are ligated to an oligonucleotide linker, and amplified by PCR using two primers. One primer is based on sequence from the instant nucleic acids, for which full length sequence is desired, and a second primer comprises sequence that hybridizes to the oligonucleotide linker to amplify the cDNA. A description of this methods is reported in WO 97/19110. A common primer may be designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends. When a single gene-specific RACE primer is paired with the common primer, preferential amplification of sequences between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook et al., 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure

herein on optional changes in amino acids to achieve altered protein structure and/or function. As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more nucleic acids of the invention can be synthesized.

EXPRESSION OF POLYPEPTIDES

The provided nucleic acid, *e.g.* a nucleic acid having a sequence of one of SEQ ID NOS:1-999), the corresponding cDNA, the polypeptide coding sequence as described above, or the full-length gene is used to express a partial or complete gene product. Constructs of nucleic acids having sequences of SEQ ID NOS:1-999 can be generated by recombinant methods, synthetically, or in a single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, *e.g.* Stemmer *et al.*, Gene (Amsterdam) (1995) 164(1):49-53.

Appropriate nucleic acid constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. The gene product encoded by a nucleic acid of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems.

The subject nucleic acid molecules are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole organism or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially.

The nucleic acids set forth in SEQ ID NOS:1-999 or their corresponding full-length nucleic acids are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters attached either at the 5' end of the sense strand or at the 3' end of the antisense strand, enhancers,

terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the nucleic acids or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

IDENTIFICATION OF FUNCTIONAL AND STRUCTURAL MOTIFS

Translations of the nucleotide sequence of the provided nucleic acids, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the nucleic acids of the invention. Also, sequences exhibiting similarity with more than one individual sequence can exhibit activities that are characteristic of either or both individual sequences.

The six possible reading frames may be translated using programs such as GCG pepdata, or GCG Frames (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.). Programs such as ORFFinder (National Center for Biotechnology Information (NCBI) a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH) http://www.ncbi.nlm.nih.gov/) may be used to identify open reading frames (ORFs) in sequences. ORF finder identifies all possible ORFs in a DNA sequence by locating the standard and alternative stop and start codons. Other ORF identification programs include Genie (Kulp *et al.* (1996).

A generalized Hidden Markov Model may be used for the recognition of genes in DNA. (ISMB-96, St. Louis, MO, AAAI/MIT Press; Reese *et al.* (1997), "Improved splice site detection in Genie". Proceedings of the First Annual International Conference on Computational Molecular Biology RECOMB 1997, Santa Fe, NM,

ACM Press, New York., P. 34.); BESTORF --Prediction of potential coding fragment in human or plant EST/mRNA sequence data using Markov Chain Models; and FGENEP -- Multiple genes structure prediction in plant genomic DNA (Solovyev *et al.* (1995) Identification of human gene structure using linear discriminant functions and dynamic programming. In Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology eds. Rawling *et al.* Cambridge, England, AAAI Press,367-375.; Solovyev *et al.* (1994) Nucl. Acids Res. **22**(24):5156-5163; Solovyev *et al,*. The prediction of human exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames, in: The Second International conference on Intelligent systems for Molecular Biology (eds. Altman et al.), AAAI Press, Menlo Park, CA (1994, 354-362) Solovyev and Lawrence, Prediction of human gene structure using dynamic programming and oligonucleotide composition, In: Abstracts of the 4th annual Keck symposium. Pittsburgh, 47,1993; Burge and Karlin (1997) J. Mol. Biol. **268**:78-94; Kulp *et al.* (1996) Proc. Conf. on Intelligent Systems in Molecular Biology '96, 134-142).

The full length sequences and fragments of the nucleic acid sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence corresponding to provided nucleic acids. Typically, a selected nucleic acid is translated in all six frames to determine the best alignment with the individual sequences. These amino acid sequences are referred to, generally, as query sequences, which are aligned with the individual sequences. Suitable databases include Genbank, EMBL, and DNA Database of Japan (DDBJ).

Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST, available by ftp at ftp://ncbi.nlm.nih.gov/.

Gapped BLAST and PSI-BLAST are useful search tools provided by NCBI. (version 2.0) (Altschul *et al.*, 1997). Position-Specific Iterated BLAST (PSI-BLAST) provides an automated, easy-to-use version of a "profile" search, which is a sensitive way to look for sequence homologues. The program first performs a gapped BLAST database search. The PSI-BLAST program uses the information from any significant alignments returned to construct a position-specific score matrix, which replaces the

query sequence for the next round of database searching. PSI-BLAST may be iterated until no new significant alignments are found. The Gapped BLAST algorithm allows gaps (deletions and insertions) to be introduced into the alignments that are returned. Allowing gaps means that similar regions are not broken into several segments. The scoring of these gapped alignments tends to reflect biological relationships more closely. The Smith-Waterman is another algorithm that produces local or global gapped sequence alignments, see Meth. Mol. Biol. (1997) 70: 173-187. Also, the GAP program using the Needleman and Wunsch global alignment method can be utilized for sequence alignments.

Results of individual and query sequence alignments can be divided into three categories, high similarity, weak similarity, and no similarity. Individual alignment results ranging from high similarity to weak similarity provide a basis for determining polypeptide activity and/or structure. Parameters for categorizing individual results include: percentage of the alignment region length where the strongest alignment is found, percent sequence identity, and e value.

The percentage of the alignment region length is calculated by counting the number of residues of the individual sequence found in the region of strongest alignment, *e.g.* contiguous region of the individual sequence that contains the greatest number of residues that are identical to the residues of the corresponding region of the aligned query sequence. This number is divided by the total residue length of the query sequence to calculate a percentage. For example, a query sequence of 20 amino acid residues might be aligned with a 20 amino acid region of an individual sequence. The individual sequence might be identical to amino acid residues 5, 9-15, and 17-19 of the query sequence. The region of strongest alignment is thus the region stretching from residue 9-19, an 11 amino acid stretch. The percentage of the alignment region length is: 11 (length of the region of strongest alignment) divided by (query sequence length) 20 or 55%.

Percent sequence identity is calculated by counting the number of amino acid matches between the query and individual sequence and dividing total number of matches by the number of residues of the individual sequences found in the region of

strongest alignment. Thus, the percent identity in the example above would be 10 matches divided by 11 amino acids, or approximately, 90.9%

E value is the probability that the alignment was produced by chance. For a single alignment, the e value can be calculated according to Karlin et al., Proc. Natl. Acad. Sci. (1990) 87:2264 and Karlin et al., Proc. Natl. Acad. Sci. (1993) 90. The e value of multiple alignments using the same query sequence can be calculated using an heuristic approach described in Altschul et al., Nat. Genet. (1994) 6:119. Alignment programs such as BLAST program can calculate the e value.

Another factor to consider for determining identity or similarity is the location of the similarity or identity. Strong local alignment can indicate similarity even if the length of alignment is short. Sequence identity scattered throughout the length of the query sequence also can indicate a similarity between the query and profile sequences. The boundaries of the region where the sequences align can be determined according to Doolittle, *supra*; BLAST or FASTA programs; or by determining the area where sequence identity is highest.

In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. The query sequence is considered to have a high similarity with a profile sequence when the p value is less than or equal to 10^{-2}. Confidence in the degree of similarity between the query sequence and the profile sequence increases as the p value become smaller.

In general, where alignment results considered to be of weak similarity, there is no minimum percent length of the alignment region nor minimum length of

alignment. A better showing of weak similarity is considered when the region of alignment is, typically, at least about 15 amino acid residues in length; more typically, at least about 20; even more typically; at least about 25 amino acid residues in length. Usually, length of the alignment region can be as much as about 30 amino acid residues; more usually, as much as about 40; even more usually, as much as about 60 amino acid residues. Further, for weak similarity, the region of alignment, typically, exhibits at least about 35% of sequence identity; more typically, at least about 40%; even more typically; at least about 45% sequence identity. Usually, percent sequence identity can be as much as about 50%; more usually, as much as about 55%; even more usually, as much as about 60%.

The query sequence is considered to have a low similarity with a profile sequence when the p value is greater than 10^{-2}. Confidence in the degree of similarity between the query sequence and the profile sequence decreases as the p values become larger.

Sequence identity alone can be used to determine similarity of a query sequence to an individual sequence and can indicate the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%. Sequence identity alone as a measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, 90 residues; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in length.

It is apparent, when studying protein sequence families, that some regions have been better conserved than others during evolution. These regions are generally important for the function of a protein and/or for the maintenance of its three- dimensional structure. By analyzing the constant and variable properties of such groups of similar sequences, it is possible to derive a signature for a protein

family or domain, which distinguishes its members from all other unrelated proteins. A pertinent analogy is the use of fingerprints by the police for identification purposes. A fingerprint is generally sufficient to identify a given individual. Similarly, a protein signature can be used to assign a new sequence to a specific family of proteins and thus to formulate hypotheses about its function. The PROSITE database is a compendium of such fingerprints (motifs) and may be used with search software such as Wisconsin GCG Motifs to find motifs or fingerprints in query sequences. PROSITE currently contains signatures specific for about a thousand protein families or domains. Each of these signatures comes with documentation providing background information on the structure and function of these proteins (Hofmann *et al.* (1999) Nucleic Acids Res. **27**:215-219; Bucher and Bairoch ., A generalized profile syntax for biomolecular sequences motifs and its function in automatic sequence interpretation (In) ISMB-94; Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology; Altman *et al.* Eds. (1994), pp 53-61, AAAI Press, Menlo Park).

Translations of the provided nucleic acids can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided nucleic acids can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene products (e.g., polypeptides) encoded by the provided nucleic acids or corresponding cDNA or genes.

Profiles can designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney et al., Nucl. Acid Res. (1996) 24(14): 2730-2739. MSAs of some protein families and motifs are available for downloading to a local server. For example, the PFAM database with MSAs of 547 different families and motifs, and the software (HMMER) to search the PFAM database may be downloaded from ftp://ftp.genetics.wustl.edu/pub/eddy/pfam-4.4/ to allow secure searches on a local server. Pfam is a database of multiple alignments of protein domains or conserved

protein regions., which represent evolutionary conserved structure that has implications for the protein's function (Sonnhammer *et al.* (1998) Nucl. Acid Res. **26**:320-322; Bateman *et al.* (1999) Nucleic Acids Res. **27**:260-262).

The 3D_ali databank (Pasarella, S. and Argos, P. (1992) Prot. Engineering 5:121-137) was constructed to incorporate new protein structural and sequence data. The databank has proved useful in many research fields such as protein sequence and structure analysis and comparison, protein folding, engineering and design and evolution. The collection enhances present protein structural knowledge by merging information from proteins of similar main-chain fold with homologous primary structures taken from large databases of all known sequences. 3D_ali databank files may be downloaded to a secure local server from http://www.embl-heidelberg.de/argos/ali/ali_form.html.

The identify and function of the gene that correlates to a nucleic acid described herein can be determined by screening the nucleic acids or their corresponding amino acid sequences against profiles of protein families. Such profiles focus on common structural motifs among proteins of each family. Publicly available profiles are known in the art.

In comparing a novel nucleic acid with known sequences, several alignment tools are available. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng et al., J. Mol. Evol. (1987) 25:351. Another method, GAP, uses the alignment method of Needleman et al., J. Mol. Biol. (1970) 48:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith *et al.* (1981) Adv. Appl. Math. **2**:482.

IDENTIFICATION OF SECRETED & MEMBRANE-BOUND POLYPEPTIDES

Secreted and membrane-bound polypeptides of the present invention are of interest. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides. A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a

polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods, Proc. Natl. Acad. Sci. USA (1981) 78:3824-3828; Kyte & Doolittle, J. Mol. Biol. (1982) 157: 105-132; and RAOAR algorithm, Degli Esposti et al., Eur. J. Biochem. (1990) 190: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the nucleic acids of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine.

IDENTIFICATION OF THE FUNCTION OF AN EXPRESSION PRODUCT

The biological function of the encoded gene product of the invention may be determined by empirical or deductive methods. One promising avenue, termed phylogenomics, exploits the use of evolutionary information to facilitate assignment of gene function. The approach is based on the idea that functional predictions can be greatly improved by focusing on how genes became similar in sequence during evolution instead of focusing on the sequence similarity itself. One of the major efficiencies that has emerged from plant genome research to date is that a large percentage of higher plant genes can be assigned some degree of function by comparing them with the sequences of genes of known function.

Alternatively, "reverse genetics" is used to identify gene function. Large collections of insertion mutants are available for *Arabidopsis*, maize, petunia, and snapdragon. These collections can be screened for an insertional inactivation of any

gene by using the polymerase chain reaction (PCR) primed with oligonucleotides based on the sequences of the target gene and the insertional mutagen. The presence of an insertion in the target gene is indicated by the presence of a PCR product. By multiplexing DNA samples, hundreds of thousands of lines can be screened and the corresponding mutant plants can be identified with relatively small effort. Analysis of the phenotype and other properties of the corresponding mutant will provide an insight into the function of the gene.

In one method of the invention, the gene function in a transgenic Arabidopsis plant is assessed with anti-sense constructs. A high degree of gene duplication is apparent in Arabidopsis, andmany of the gene duplications in Arabidopsis are very tightly linked. Large numbers of transgenic Arabidopsis plants can be generated by infecting flowers with Agrobacterium tumefaciens containing an insertional mutagen, a method of gene silencing based on producing double-stranded RNA from bidirectional transcription of genes in transgenic plants can be broadly useful for high-throughput gene inactivation (Clough and Bent (1999) Plant J. **17**; Waterhouse *et al.* (1998) Proc. Natl. Acad. Sci. U.S.A. **95**:13959). This method may use promoters that are expressed in only a few cell types or at a particular developmental stage or in response to an external stimulus. This could significantly obviate problems associated with the lethality of some mutations.

Virus-induced gene silencing may also find use for suppressing gene function. This method exploits the fact that some or all plants have a surveillance system that can specifically recognize viral nucleic acids and mount a sequence-specific suppression of viral RNA accumulation. By inoculating plants with a recombinant virus containing part of a plant gene, it is possible to rapidly silence the endogenous plant gene.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense nucleic acids based on a selected nucleic acid sequence can interfere with expression of the corresponding gene. Antisense nucleic acids are typically generated within the cell by expression from antisense constructs that contain the antisense strand as the

transcribed strand. Antisense nucleic acids based on the disclosed nucleic acids will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense nucleic acid. The expression products of control cells and cells treated with the antisense construct are compared to detect the protein product of the gene corresponding to the nucleic acid upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

As an alternative method for identifying function of the gene corresponding to a nucleic acid disclosed herein, dominant negative mutations are readily generated for corresponding proteins that are active as homomultimers. A mutant polypeptide will interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants (see for example, Herskowitz (1987) Nature 329:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

Another approach for discovering the function of genes utilizes gene chips and microarrays. DNA sequences representing all the genes in an organism can be placed on miniature solid supports and used as hybridization substrates to quantitate the expression of all the genes represented in a complex mRNA sample. This information is used to provide extensive databases of quantitative information about the degree to which each gene responds to pathogens, pests, drought, cold, salt, photoperiod, and other environmental variation. Similarly, one obtains extensive information about which genes respond to changes in developmental processes such as germination and flowering. One can therefore determine which genes respond to the phytohormones, growth regulators, safeners, herbicides, and related agrichemicals. These databases of gene expression information provide insights into the "pathways" of genes that control complex responses. The accumulation of DNA

microarray or gene chip data from many different experiments creates a powerful opportunity to assign functional information to genes of otherwise unknown function. The conceptual basis of the approach is that genes that contribute to the same biological process will exhibit similar patterns of expression. Thus, by clustering genes based on the similarity of their relative levels of expression in response to diverse stimuli or developmental or environmental conditions, it is possible to assign functions to many genes based on the known function of other genes in the cluster.

CONSTRUCTION OF POLYPEPTIDES OF THE INVENTION AND VARIANTS THEREOF

The polypeptides of the invention include those encoded by the disclosed nucleic acids. These polypeptides can also be encoded by nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed nucleic acids. Thus, the invention includes within its scope a polypeptide encoded by a nucleic acid having the sequence of any one of SEQ ID NOS: 1-999 or a variant thereof.

In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited nucleic acid, the polypeptide encoded by the gene represented by the recited nucleic acid, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein. In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide of the invention, as measured by BLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, i.e., the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, *e.g.* are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition

that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid substituted.

Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 amino acids (aa) to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a nucleic acid having a sequence of any SEQ ID NOS:1-999, or a homolog thereof.

The protein variants described herein are encoded by nucleic acids that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

LIBRARIES AND ARRAYS

In general, a library of biopolymers is a collection of sequence information, which information is provided in either biochemical form (e.g., as a collection of nucleic acid or polypeptide molecules), or in electronic form (e.g., as a collection of

genetic sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The term biopolymer, as used herein, is intended to refer to polypeptides, nucleic acids, and derivatives thereof, which molecules are characterized by the possession of genetic sequences either corresponding to, or encoded by, the sequences set forth in the provided sequence list (seqlist). The sequence information can be used in a variety of ways, e.g., as a resource for gene discovery, as a representation of sequences expressed in a selected cell type, *e.g.* cell type markers, *etc.*

The nucleic acid libraries of the subject invention include sequence information of a plurality of nucleic acid sequences, where at least one of the nucleic acids has a sequence of any of SEQ ID NOS:1-999. By plurality is meant one or more, usually at least 2 and can include up to all of SEQ ID NOS:1-999. The length and number of nucleic acids in the library will vary with the nature of the library, e.g., if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the sequences or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, e.g. the nucleic acid sequences of any of the nucleic acids of SEQ ID NOS:1-999, can be recorded on computer readable media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium,

using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, etc. In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (e.g., searchable files, executable files, etc, including, but not limited to, for example, search program software, etc.)

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the BLAST (Altschul et al., supra.) and BLAZE (Brutlag et al. Comp. Chem. (1993) 17:203) search algorithms on a Sybase system can be used identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif with the stored sequence information. Search means are used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, e.g. MacPattern (EMBL), BLASTN, BLASTX (NCBI) and tBLASTX. A "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more

amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues.

A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but arc not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks fragments of the genome possessing varying degrees of homology to a target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences and identifies the degree of sequence similarity contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention.

As discussed above, the "library" of the invention also encompasses biochemical libraries of the nucleic acids of SEQ ID NOS:1-999, e.g., collections of nucleic acids representing the provided nucleic acids. The biochemical libraries can take a variety of forms, e.g. a solution of cDNAs, a pattern of probe nucleic acids stably bound to a surface of a solid support (microarray) and the like. By array is meant an article of manufacture that has a solid support or substrate with one or more nucleic acid targets on one of its surfaces, where the number of distinct nucleic may be in the hundreds, thousand, or tens of thousands. Each nucleic acid will comprise at 18 nt and often at least 25 nt, and often at least 100 to 1000 nucleotides, and may represent up to a complete coding sequence or cDNA.. A variety of

different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by SEQ ID NOS:1-999.

GENETICALLY ALTERED CELLS AND TRANSGENICS

The subject nucleic acids can be used to create genetically modified and transgenic organisms, usually plant cells and plants, which may be monocots or dicots. The term transgenic, as used herein, is defined as an organism into which an exogenous nucleic acid construct has been introduced, generally the exogenous sequences are stably maintained in the genome of the organism. Of particular interest are transgenic organisms where the genomic sequence of germ line cells has been stably altered by introduction of an exogenous construct.

Typically, the transgenic organism is altered in the genetic expression of the introduced nucleotide sequences as compared to the wild-type, or unaltered organism. For example, constructs that provide for over-expression of a targeted sequence, sometimes referred to as a "knock-in", provide for increased levels of the gene product. Alternatively, expression of the targeted sequence can be down-regulated or substantially eliminated by introduction of a "knock-out" construct, which may direct transcription of an anti-sense RNA that blocks expression of the naturally occurring mRNA, by deletion of the genomic copy of the targeted sequence, *etc.*

In one method, large numbers of genes are simultaneously introduced in order to explore the genetic basis of complex traits, for example by making plant artificial chromosome (PLAC) libraries. The centromeres in *Arabidopsis* have been mapped and current genome sequencing efforts will extend through these regions. Because *Arabidopsis* telomeres are very similar to those in yeast one may use a hybrid sequence of alternating plant and yeast sequences that function in both types of organisms, developing yeast artificial chromosome-PLAC libraries, and then

introducing them into a suitable plant host to evaluate the phenotypic consequences. By providing a defined chromosomal environment for cloned genes, the use of PLACs may also enhance the ability to produce transgenic plants with defined levels of gene expression.

It has been found in many organisms that there is significant redundancy in the representation of genes in a genome. That is, a particular gene function is likely by represented by multiple copies of similar coding sequences in the genome. These copies are typically conserved in the amino acid sequence, but may diverge in the sequence of non-translated sequences, and in their codon usage. In order to knock out a particular genetic function in an organism, it may not be sufficient to delete a genomic copy of a single gene. In such cases it may be preferable to achieve a genetic knock-out with an anti-sense construct, particularly where the sequence is aligned with the coding portion of the mRNA.

Methods of transforming plant cells are well-known in the art, and include protoplast transformation, tungsten whiskers (Coffee et al., U.S. Pat. No. 5,302,523, issued Apr. 12, 1994), directly by microorganisms with infectious plasmids, use of transposons (U.S. Patent No. 5,792,294), infectious viruses, the use of liposomes, microinjection by mechanical or laser beam methods, by whole chromosomes or chromosome fragments, electroporation, silicon carbide fibers, and microprojectile bombardment.

For example, one may utilize the biolistic bombardment of meristem tissue, at a very early stage of development, and the selective enhancement of transgenic sectors toward genetic homogeneity, in cell layers that contribute to germline transmission. Biolistics-mediated production of fertile, transgenic maize is described in Gordon-Kamm *et al.* (1990), Plant Cell **2**:603; Fromm *et al.* (1990) Bio/Technology **8**: 833, for example. Alternatively, one may use a microorganism, including but not limited to, *Agrobacterium tumefaciens* as a vector for transforming the cells, particularly where the targeted plant is a dicotyledonous species. See, for example, U.S. Patent No. 5,635,381. Leung *et al.* (1990) Curr. Genet. **17**(5):409-11 describe integrative transformation of three fertile hermaphroditic strains of *Arabidopsis*

thaliana using plasmids and cosmids that contain an *E. coli* gene linked to *Aspergillus nidulans* regulatory sequences.

Preferred expression cassettes for cereals may include promoters that are known to express exogenous DNAs in corn cells. For example, the Adhl promoter has been shown to be strongly expressed in callus tissue, root tips, and developing kernels in corn. Promoters that are used to express genes in corn include, but are not limited to, a plant promoter such as the, CaMV 35S promoter (Odell et al., Nature, 313, 810 (1985)), or others such as CaMV 19S (Lawton et al., Plant Mol. Biol., 9, 31F (1987)), nos (Ebert et al., PNAS USA, 84, 5745 (1987)), Adh (Walker et al., PNAS USA, 84, 6624 (1987)), sucrose synthase (Yang et al., PNAS USA, 87, 4144 (1990)), .alpha.-tubulin, ubiquitin, actin (Wang et al., Mol. Cell. Biol., 12, 3399 (1992)), cab (Sullivan et al., Mol. Gen. Genet, 215, 431 (1989)), PEPCase (Hudspeth et al., Plant Mol. Biol., 12, 579 (1989)), or those associated with the R gene complex (Chandler et al., The Plant Cell, 1, 1175 (1989)). Other promoters useful in the practice of the invention are known to those of skill in the art.

Tissue-specific promoters, including but not limited to, root-cell promoters (Conkling et al., Plant Physiol., 93, 1203 (1990)), and tissue-specific enhancers (Fromm et al., The Plant Cell, 1, 977 (1989)) are also contemplated to be particularly useful, as are inducible promoters such as water-stress-, ABA- and turgor-inducible promoters (Guerrero et al., Plant Molecular Biology, 15, 11-26)), and the like.

Regulating and/or limiting the expression in specific tissues may be functionally accomplished by introducing a constitutively expressed gene (all tissues) in combination with an antisense gene that is expressed only in those tissues where the gene product is not desired. Expression of an antisense transcript of this preselected DNA segment in an rice grain, using, for example, a zein promoter, would prevent accumulation of the gene product in seed. Hence the protein encoded by the preselected DNA would be present in all tissues except the kernel.

Alternatively, one may wish to obtain novel tissue-specific promoter sequences for use in accordance with the present invention. To achieve this, one may first isolate cDNA clones from the tissue concerned and identify those clones which are expressed specifically in that tissue, for example, using Northern blotting or

DNA microarrays. Ideally, one would like to identify a gene that is not present in a high copy number, but which gene product is relatively abundant in specific tissues. The promoter and control elements of corresponding genomic clones may then be localized using the techniques of molecular biology known to those of skill in the art. Alternatively, promoter elements can be identified using enhancer traps based on T-DNA and/or transposon vector systems (see, for example, Campisi *et al.* (1999) Plant J. **17**:699-707; Gu *et al.* (1998) Development **125**:1509-1517).

In some embodiments of the present invention expression of a DNA segment in a transgenic plant will occur only in a certain time period during the development of the plant. Developmental timing is frequently correlated with tissue specific gene expression. For example, in corn expression of zein storage proteins is initiated in the endosperm about 15 days after pollination.

Ultimately, the most desirable DNA segments for introduction into a plant genome may be homologous genes or gene families which encode a desired trait (e.g., increased disease resistance) and which are introduced under the control of novel promoters or enhancers, etc., or perhaps even homologous or tissue-specific (e.g., root-, grain- or leaf-specific) promoters or control elements.

The genetically modified cells are screened for the presence of the introduced genetic material. The cells may be used in functional studies, drug screening, *etc.*, *e.g.* to study chemical mode of action, to determine the effect of a candidate agent on pathogen growth, infection of plant cells, *etc.*

The modified cells are useful in the study of genetic function and regulation, for alteration of the cellular metabolism, and for screening compounds that may affect the biological function of the gene or gene product. For example, a series of small deletions and/or substitutions may be made in the host's native gene to determine the role of different domains and motifs in the biological function. Specific constructs of interest include anti-sense, as previously described, which will reduce or abolish expression, expression of dominant negative mutations, and over-expression of genes.

Where a sequence is introduced, the introduced sequence may be either a complete or partial sequence of a gene native to the host, or may be a complete or

partial sequence that is exogenous to the host organism, *e.g.*, an *A. thaliana* sequence inserted into wheat plants. A detectable marker, such as aldA, *lac Z*, *etc.* may be introduced into the locus of interest, where upregulation of expression will result in an easily detected change in phenotype.

One may also provide for expression of the gene or variants thereof in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues, or at abnormal times of development, during sporulation, *etc.* By providing expression of the protein in cells in which it is not normally produced, one can induce changes in cell behavior.

DNA constructs for homologous recombination will comprise at least a portion of the provided gene or of a gene native to the species of the host organism, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus (see Kempin *et al.* (1997) Nature **389**:802-803). DNA constructs for random integration or episomal maintenance need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art.

Embodiments of the invention provide processes for enhancing or inhibiting synthesis of a protein in a plant by introducing a provided nucleic acids sequence into a plant cell, where the nucleic acid comprises sequences encoding a protein of interest. For example, enhanced resistance to pathogens may be achieved by inserting a nucleic acid encoding an activator in a vector downstream from a promoter sequence capable of driving constitutive high-level expression in a plant cell. When grown into plants, the transgenic plants exhibit increased synthesis of resistance proteins, and increased resistance to pathogens.

Other embodiments of the invention provide processes for enhancing or inhibiting synthesis of a tolerance factor in a plant by introducing a nucleic acid of the invention into a plant cell, where the nucleic acid comprises sequences encoding a tolerance factor. For example, enhanced tolerance to an environmental stress may be achieved by inserting a nucleic acid encoding an activator in a vector downstream from a promoter sequence capable of driving constitutive high-level expression in a

plant cell. When grown into plants, the transgenic plants exhibit increased synthesis of tolerance proteins, and increased tolerance to environmental stress.

Factors which are involved, directly or indirectly in biosynthetic pathways whose products are of commercial, nutritional, or medicinal value include any factor, usually a protein or peptide, which regulates such a biosynthetic pathway (e.g., an activator or repressor); which is an intermediate in such a biosynthetic pathway; or which is a product that increases the nutritional value of a food product; a medicinal product; or any product of commercial value and/or research interest. Plant and other cells may be genetically modified to enhance a trait of interest, by upregulating or down-regulating factors in a biosynthetic pathway.

SCREENING ASSAYS

The polypeptides encoded by the provided nucleic acid sequences, and cells genetically altered to express such sequences, are useful in a variety of screening assays to determine effect of candidate inhibitors, activators., or modifiers of the gene product. One may determine what insecticides, fungicides and the like have an enhancing or synergistic activity with a gene. Alternatively, one may screen for compounds that mimic the activity of the protein. Similarly, the effect of activating agents may be used to screen for compounds that mimic or enhance the activation of proteins. Candidate inhibitors of a particular gene product are screened by detecting decreased from the targeted gene product.

The screening assays may use purified target macromolecules to screen large compound libraries for inhibitory drugs; or the purified target molecule may be used for a rational drug design program, which requires first determining the structure of the macromolecular target or the structure of the macromolecular target in association with its customary substrate or ligand. This information is then used to design compounds which must be synthesized and tested further. Test results are used to refine the molecular models and drug design process in an iterative fashion until a lead compound emerges.

Drug screening may be performed using an *in vitro* model, a genetically altered cell, or purified protein. One can identify ligands or substrates that bind to,

modulate or mimic the action of the target genetic sequence or its product. A wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. The purified protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions.

Where the nucleic acid encodes a factor involved in a biosynthetic pathway, as described above, it may be desirable to identify factors, *e.g.*, protein factors, which interact with such factors. One can identify interacting factors, ligands, substrates that bind to, modulate or mimic the action of the target genetic sequence or its product. A wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. *In vivo* assays for protein-protein interactions in *E. coli* and yeast cells are also well-established (see Hu *et al.* (2000) Methods **20**:80-94; and Bai and Elledge (1997) Methods Enzymol. **283**:141-156).

The purified protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions. It may also be of interest to identify agents that modulate the interaction of a factor identified as described above with a factor encoded by a nucleic acid of the invention. Drug screening can be performed to identify such agents. For example, a labeled in vitro protein-protein binding assay can be used, which is conducted in the presence and absence of an agent being tested.

The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of altering or mimicking a physiological function. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents

comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and organism extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles, *e.g.* magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin *etc.* For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as

protease inhibitors, nuclease inhibitors, anti-microbial agents, *etc.* may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40° C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

The compounds having the desired biological activity may be administered in an acceptable carrier to a host. The active agents may be administered in a variety of ways. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.01-100 wt.%.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a complex" includes a plurality of such complexes and reference to "the formulation" includes reference to one or more formulations and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the methods and methodologies that are described in the publications which might be used in connection with the presently described invention. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the

subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (*e.g.* amounts, temperature, concentrations, *etc.*) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric.

EXPERIMENTAL

Cloning and Characterization of *Arabidopsis thaliana* Genes.

Following DNA isolation, sequencing was performed using the Dye Primer Sequencing protocol, below. The sequencing reactions were loaded by hand onto a 48 lane ABI 377 and run on a 36 cm gel with the 36E-2400 run module and extraction. Gel analysis was performed with ABI software.

The Phred program was used to read the sequence trace from the ABI sequencer, call the bases and produce a sequence read and a quality score for each base call in the sequence., (Ewing *et al.* (1998) Genome Research **8**:175-185; Ewing and Green (1998) Genome Research **8**:186-194.) PolyPhred may be used to detect single nucleotide polymorphisms in sequences (Kwok *et al.* (1994) Genomics **25**:615-622; Nickerson *et al.* (1997) Nucleic Acids Research **25**(14):2745-2751.)

MicroWave Plasmid Protocol: Fill Beckman 96 deep-well growth blocks with 1 ml of TB containing 50 μg of ampicillin per ml. Inoculate each well with a colony picked with a toothpick or a 96-pin tool from a glycerol stock plate. Cover the blocks with a plastic lid and tape at two ends to hold lid in place. Incubate overnight (16-24 hours depending on the host stain) at 37° C with shaking at 275 rpm in a New Brunswick platform shaker. Pellet cells by centrifugation for 20 minutes at 3250 rpm in a Beckman GS-R6K, decant TB and freeze pelleted cell in the 96 well block. Thaw blocks on the bench when ready to continue.

Prepare the MW-Tween20 solution

For four blocks:	For 16 blocks:
50ml STET/TWEEN20	200ml STET/TWEEN
2 tubes RNAse (10mg/ml,600ulea)	8 tubes RNAse
1 tube lysozyme (25mg)	4 tubes lysozyme

Pipette RNAse and Lysozyme into the corner of a beaker. Add Tween 20 solution and swirl to mix completely. Use the Multidrop (or Biohit) to add 25ul of sterile H_20 (from the L size autoclaved bottles) to each well. Resuspend the pellets by vortexing on setting 10 of the platform vortexer. Check pellets after 4 min. and repeat as necessary to resuspend completely. Use the multidrop to add 70 μl of the freshly prepared MW-Tween 20 solution to each well. Vortex at setting 6 on the platform vortex for 15 seconds. Do not cause frothing.

Incubate the blocks at room temperature for 5 min. Place two blocks at a time in the microwave (1000 Watts) with the tape (placed on the H1 to H12 side of the block) facing away from each other and turn on at full power for 30 seconds. Rotate the blocks so that the tapes face towards each other and turn on at full power again for 30 seconds.

Immediately remove the blocks from the microwave and add 300 μl of sterile ice cold H_2O with the Multidrop. Seal the blocks with foil tape and place them in an H_20/ice bath.

Vortex the blocks on 5 for 15 seconds and leave them in the H_2O/Ice bath. Return to step 7 until all the blocks are in the ice water bath. Incubate the blocks for 15 minutes on ice. Spin the blocks for 30 minutes in the Beckman GS-6KR with GH3.8 rotor with Microplus carrier at 3250rpm.

Transfer 100 µl of the supernatant to Corning/Costar round bottom 96 well trays. Cover with foil and put into fridge if to be sequenced right away. If not to be sequenced in the next day, freeze them at –20° C.

Dye Primer Sequencing: Spin down the DP brew trays and DNA template by pulsing in the Beckman GS-6KR with GH3.8 rotor with Microplus carrier. Big Dye Primer reaction mix trays (one 96 well cycleplate (Robbins) for each nucleotide), 3 microliters of reaction mix per well.

Use twelve channel pipetter (Costar) to add 2 μl of template to one each G,A,T,C, trays for each template plate. Pulse again to get both the reaction mix and template into the bottom of the cycle plate and put them into the MJ Research DNA Tetrad (PTC-225).

Start program Dye-Primer. Dye-primer is:

96° C, 1 min 1 cycle

96° C, 10 sec.

55° C, 5 sec.

70° C, 1 min 15 cycles

96° C, 10 sec.

70° C, 1 min. 15 cycles

4° C soak

When done cycling, using the Robbins Hydra 290 add 100 μl of 100 % ethanol to the A reaction cycle plate and pool the contents of all four cycle plates into the appropriate well.

To perform ethanol precipitation: Use Hydra program 4 to add 100 μl 100% ethanol to each A tray. Use Hydra program 5 to transfer the ethanol and therefore combine the samples from plate to plate. Once the G, A, T, and C trays of each block are mixed, spin for 30 minutes at 3250 in the Beckman. Pour off the ethanol with a firm shake and blot on a paper towel before drying in the speed vac (~10 minutes or until dry). If ready to load add 3 μl dye and denature in the oven at 95° C for ~5 minutes and load 2 μl. If to store, cover with tape and store at –20°C.

Common Solutions

Terrific Broth

Per liter:

900 ml H_2O

12 g bacto tryptone

24 g bacto-yeast extract

4 ml glycerol

Shake until dissolved and then autoclave. Allow the solution to cool to 60° C or less and then add 100 ml of sterile 0.17M KH_2PO_4, 0.72M K_2HPO_4 (in the hood w/ sterile technique).

0.17M KH_2PO_4, 0.72M K_2HPO_4

Dissolve 2.31g of KH_2PO_4 and 12.54g of K_2HPO_4 in 90 ml of H_2O.

Adjust volume to 100 ml with H_2O and autoclave.

Sequence loading Dye

20 ml deionized formamide

3.6 ml dH_20

400 µl 0.5M EDTA, pH 8.0

0.2 g Blue Dextran

*Light sensitive, cover in foil or store in the dark.

STET/TWEEN

10 ml 5M NaCl

5 ml 1M Tris, pH 8.0

1 ml 0.5M EDTA., pH 8.0

25ml Tween20

Bring volume to 500 ml with H_2O

The sequencing reactions are run on an ABI 377 sequencer per manufacturer's' instructions. The sequencing information obtained each run are analyzed as follows.

Sequencing reads are screened for ribosomal., mitochondrial., chloroplast or human sequence contamination.. In good sequences, vector is marked by x's. These sequences go into biolims regardless of whether or not they pass the criteria for a 'good' sequence. This criteria is >= 100 bases with phred score of >=20 and 15 of these bases adjacent to each other.

Sequencing reads that pass the criteria for good sequences are downloaded for assembly into consensus sequences (contigs). The program Phrap (copyrighted by Phil Green at University of Washington, Seattle, WA) utilizes both the Phred sequence information and the quality calls to assemble the sequencing reads. Parameters used with Phrap were determined empirically to minimize assembly of

chimeric sequences and maximize differential detection of closely related members of gene families. The following parameters were used with the Phrap program to perform the assembly:

Penalty	–6	Penalty for mismatches(substitutions)
Minmatch	40	Minimum length of matching sequence to use in assembly of reads
Trim penalty	0	penalty used for identifying degenerate sequence at beginning and end of read.
Minscore	80	Minimum alignment score

Results from the Phrap analysis yield either contigs consisting of a consensus of two or more overlapping sequence reads, or singlets that are non-overlapping .

The contig and singlets assembly were further analyzed to eliminate low quality sequence utilizing a program to filter sequences based on quality scores generated by the Phred program. The threshold quality for "high quality" base calls is 20. Sequences with less than 50 contiguous high quality bases calls at the beginning of the sequence, and also at the end of the sequence were discarded. Additionally, the maximum allowable percentage of "low quality base calls in the final sequence is 2%, otherwise the sequence is discarded.

The stand-alone BLAST programs and Genbank databases were downloaded from NCBI for use on secure servers at the Paradigm Genetics, Inc. site. The sequences from the assembly were compared to the GenBank NR database downloaded from NCBI using the gapped version (2.0) of BLASTX. BLASTX translates the DNA sequence in all six reading frames and compares it to an amino acid database. Low complexity sequences are filtered in the query sequence. (Altschul *et al.* (1997) Nucleic Acids Res **25**(17):3389-402).

Genbank sequences found in the BLASTX search with an E Value of less than $1e^{-10}$ are considered to be highly similar, and the Genbank definition lines were used to annotate the query sequences.

When no significantly similar sequences were found as a result of the BLASTX search, the query sequences were compared with the PROSITE database (Bairoch,

A. (1992) PROSITE: A dictionary of sites and patterns in proteins. Nucleic Acids Research 20:2013-2018.) to locate functional motifs.

Query sequences were first translated in six reading frames using the Wisconsin GCG pepdata program (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.). The Wisconsin GCG motifs program (Wisconsin Package Version 10.0, Genetics Computer Group (GCG) , Madison, Wisconsin, USA.) was used to locate motifs in the peptide sequence, with no mismatches allowed. Motif names from the PROSITE results were used to annotate these query sequences.

Table 1

SEQ ID	Reference	Annotation
1	2028001	Tyr_Phospho_Site(512-519)
2	2028002	1E-30 >gi|4220454 (AC006216) Similar to gi|3413714 T19L18.21 myrosinase-binding protein from Arabidopsis thaliana BAC gb|AC004747. ESTs gb|65870 and gb|T20812 come from this gene. [Arabidopsis thaliana] Length = 303
3	2028003	1E-133 >sp|P43297|RD21_ARATH CYSTEINE PROTEINASE RD21A PRECURSOR >gi|541857|pir||JN0719 drought-inducible cysteine proteinase (EC 3.4.22.-) RD21A precursor - Arabidopsis thaliana >gi|435619|dbj|BAA02374| (D13043) thiol protease [Arabidopsis thaliana] Length = 462
4	2028004	5E-60 >gb|AAD56998.1|AC009465_12 (AC009465) mitogen activated protein kinase kinase [Arabidopsis thaliana] Length = 700
5	2028005	1E-28 >gb|AAD36643.1|AE001802_12 (AE001802) hemolysin [Thermotoga maritima] Length = 267
6	2028006	4E-41 >emb|CAA72903| (Y12227) topoisomerase [Arabidopsis thaliana] Length = 618
7	2028007	1E-103 >emb|CAB36783.1| (AL035525) aminopeptidase-like protein [Arabidopsis thaliana] Length = 873
8	2028008	2E-26 >sp|P46810|GUAA_MYCLE GMP SYNTHASE [GLUTAMINE-HYDROLYZING] (GLUTAMINE AMIDOTRANSFERASE) (GMP SYNTHETASE) >gi|2145847|pir||S72813 GMP synthase (glutamine-hydrolysing) (EC 6.3.5.2) guaA - Mycobacterium leprae >gi|466934 (U00015) guaA; B1620_C2_205 [Mycobacterium leprae] Length = 590
9	2028009	Tyr_Phospho_Site(706-713)
10	2028010	2E-33 >gb|AAD42941.1|AF091621_1 (AF091621) ubiquitin-conjugating enzyme E2 [Catharanthus roseus] Length = 153
11	2028011	1E-14 >gi|2829899 (AC002311) similar to ripening-induced protein, gp|AJ001449|2465015 and major#latex protein, gp|X91961|1107495 [Arabidopsis thaliana] Length = 160
12	2028012	Tyr_Phospho_Site(900-908)
13	2028013	2E-37 >emb|CAB52246.1| (AJ245478) alpha galactosyltransferase [Trigonella foenum-graecum] Length = 438
14	2028014	Tyr_Phospho_Site(181-187)
15	2028015	Rgd(201-203)
16	2028016	3E-70 >sp|Q08770|RL10_ARATH 60S RIBOSOMAL PROTEIN L10 (WILM'S

		TUMOR SUPPRESSOR PROTEIN HOMOLOG) >gi|478401|pir||JQ2244 ribosomal protein L10.e, cytosolic - Arabidopsis thaliana >gi|17682|emb|CAA78856| (Z15157) Wilm's tumor suppressor homologue [Arabidopsis thaliana] Length = 220
17	2028017	1E-80 >gi|2924779 (AC002334) 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] >gi|2981616|dbj|BAA25248| (AB008854) 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] >gi|2981618|dbj|BAA25249| (AB008855) 3-ketoacyl-CoA thiolase [Arabidopsis thaliana] Length = 462
18	2028018	3' Tyr_Phospho_Site(224-232)
19	2028019	3' Pkc_Phospho_Site(35-37)
20	2028020	5' Pkc_Phospho_Site(86-88)
21	2028021	5' 3E-21 >gi|3123745|dbj|BAA25999| (AB013447) aluminum-induced [Brassica napus] Length = 244
22	2028022	5' Tyr_Phospho_Site(211-218)
23	2028023	5' 3E-32 >gi|461812|sp|Q05047|CP72_CATRO CYTOCHROME P450 72A1 (CYPLXXII) (PROBABLE GERANIOL-10-HYDROXYLASE) (GE10H) >gi|167484 (L10081) Cytochrome P-450 protein [Catharanthus roseus] >gi|445604|prf||1909351A cytochrome P450 [Catharanthus roseus] Length = 524
24	2028024	5' Tyr_Phospho_Site(825-833)
25	2028025	5' 2E-75 >gi|4006827|gb|AAC95169.1| (AC005970) subtilisin-like protease [Arabidopsis thaliana] Length = 754
26	2028026	5E-40 >gi|135915|sp|P28493|PR5_ARATH PATHOGENESIS-RELATED PROTEIN 5 PRECURSOR (PR-5) >gi|322559|pir||JQ1695 pathogenesis-related protein 5 precursor - Arabidopsis thaliana >gi|166865 (M90510) thaumatin-like protein [Arabidopsis thaliana] >gi|1448919 (L78079) thaumatin-like protein [Arabidop
27	2028027	8E-24 >gb|AAD15390| (AC006223) sugar starvation-induced protein [Arabidopsis thaliana] Length = 256
28	2028028	9E-34 >sp|Q39230|SYS_ARATH SERYL-TRNA SYNTHETASE (SERINE—TRNA LIGASE) (SERRS) >gi|2129737|pir||S71293 seryl-tRNA synthetase - Arabidopsis thaliana >gi|1359497|emb|CAA94388| (Z70313) seryl-tRNA Synthetase [Arabidopsis thaliana] Length = 451
29	2028029	4E-57 >sp|P21528|MDHC_PEA MALATE DEHYDROGENASE [NADP], CHLOROPLAST PRECURSOR (NADP-MDH) >gi|481222|pir||S38346 malate dehydrogenase (NADP+) (EC 1.1.1.82) - garden pea >gi|397475|emb|CAA52614| (X74507) malate dehydrogenase (NADP+) [Pisum sativum] Length = 441
30	2028030	Rgd(1079-1081)
31	2028031	Tyr_Phospho_Site(722-728)
32	2028032	3E-23 >emb|CAB10154| (Z97211) probable involvement in ergosterol synthesis [Schizosaccharomyces pombe] Length = 1213
33	2028033	1E-102 >dbj|BAA28531| (D78598) cytochrome P450 monooxygenase [Arabidopsis thaliana] >gi|5262761|emb|CAB45909.1| (AL080283) cytochrome P450 monooxygenase [Arabidopsis thaliana] Length = 499
34	2028034	5E-36 >sp|Q42885|ARC2_LYCES CHORISMATE SYNTHASE 2 PRECURSOR (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE PHOSPHOLYASE 2) >gi|542027|pir||S40409 chorismate synthase (EC 4.6.1.4) 2 precursor - tomato >gi|410484|emb|CAA79854| (Z21791) chorismate synthase 2 [Lycopersicon esculentum] Length = 431
35	2028035	Tyr_Phospho_Site(19-25)
36	2028036	1E-123 >emb|CAA19688.1| (AL024486) aspartate kinase-homoserine dehydrogenase-like protein [Arabidopsis thaliana] Length = 916
37	2028037	2E-23 >gb|AAD48585.1| (AF110645) candidate tumor suppressor p33 ING1 homolog [Homo sapiens] Length = 249
38	2028038	Tyr_Phospho_Site(939-945)

39	2028039	1E-49 >gi|1619956 (U72151) voltage-gated chloride channel [Arabidopsis thaliana] Length = 773
40	2028040	1E-22 >gi|2338712 (AF013959) metallothionein-like protein [Arabidopsis thaliana] Length = 69
41	2028041	Pkc_Phospho_Site(45-47)
42	2028042	5E-49 >sp|Q05047|CP72_CATRO CYTOCHROME P450 72A1 (CYPLXXII) (PROBABLE GERANIOL-10-HYDROXYLASE) (GE10H) >gi|167484 (L10081) Cytochrome P-450 protein [Catharanthus roseus] >gi|445604|prf||1909351A cytochrome P450 [Catharanthus roseus] Length = 524
43	2028043	Pkc_Phospho_Site(62-64)
44	2028044	3' 1E-34 >gi|6016708|gb|AAF01534.1|AC009325_4 (AC009325) protein kinase [Arabidopsis thaliana] Length = 411
45	2028045	3' Tyr_Phospho_Site(46-53)
46	2028046	3' Tyr_Phospho_Site(297-304)
47	2028047	3' Tyr_Phospho_Site(675-683)
48	2028048	3' Tyr_Phospho_Site(77-84)
49	2028049	3' Tyr_Phospho_Site(734-740)
50	2028050	5' 3E-76 >gi|2864613|emb|CAA16960| (AL021811) S-receptor kinase -like protein [Arabidopsis thaliana] >gi|4049333|emb|CAA22558.1| (AL034567) S-receptor kinase-like protein [Arabidopsis thaliana] Length = 778
51	2028051	5' 3E-48 >gi|1514643|emb|CAA94437| (Z70524) PDR5-like ABC transporter [Spirodela polyrrhiza] Length = 1441
52	2028052	Pkc_Phospho_Site(18-20)
53	2028053	7E-30 >sp|O07051|LTAA_AERJA L-ALLO-THREONINE ALDOLASE (L-ALLO-TA) (L-ALLO-THREONINE ACETALDEHYDE-LYASE) >gi|2190272|dbj|BAA20404| (D87890) L-allo-threonine aldolase [Aeromonas jandaei] Length = 338
54	2028054	8E-68 >gb|AAD46410.1|AF096260_1 (AF096260) ER66 protein [Lycopersicon esculentum] Length = 558
55	2028055	9E-60 >dbj|BAA74589| (AB021934) nicotianamine synthase [Arabidopsis thaliana] Length = 320
56	2028056	3E-67 >gi|2281645 (AF003103) AP2 domain containing protein RAP2.10 [Arabidopsis thaliana] >gi|2632063|emb|CAA05630.1| (AJ002598) TINY-like protein [Arabidopsis thaliana] Length = 259
57	2028057	Tyr_Phospho_Site(473-481)
58	2028058	Tyr_Phospho_Site(216-223)
59	2028059	4E-59 >sp|P34881|MTDM_ARATH DNA (CYTOSINE-5)-METHYLTRANSFERASE (DNA METHYLTRANSFERASE) (DNA METASE) >gi|1363480|pir||S59604 DNA (cytosine-5-)-methyltransferase (EC 2.1.1.37) - Arabidopsis thaliana >gi|304107 (L10692) cytosine-5 methyltransferase [Arabidopsis thaliana] Length = 1534
60	2028060	Tyr_Phospho_Site(426-434)
61	2028061	3' Pkc_Phospho_Site(147-149)
62	2028062	3' 7E-28 >gi|1702872|emb|CAA70862| (Y09667) ferredoxin-dependent glutamate synthase [Arabidopsis thaliana] Length = 1648
63	2028063	3' Pkc_Phospho_Site(4-6)
64	2028064	3' 2E-81 >gi|4006882|emb|CAB16800.1| (Z99707) UDP-glucuronyltransferase-like protein [Arabidopsis thaliana] Length = 544
65	2028065	3' Pkc_Phospho_Site(48-50)
66	2028066	5' Tyr_Phospho_Site(696-704)
67	2028067	5' 3E-76 >gi|3738320 (AC005170) cinnamoyl CoA reductase [Arabidopsis thaliana] Length = 303
68	2028068	5' Rgd(11-13)

69	2028069	5' 3E-74 >gi|134103|sp|P21240|RUBB_ARATH RUBISCO SUBUNIT BINDING-PROTEIN BETA SUBUNIT PRECURSOR (60 KD CHAPERONIN BETA SUBUNIT) (CPN-60 BETA) Length = 600
70	2028070	5' Tyr_Phospho_Site(407-414)
71	2028071	5' 2E-15 >gi|3157926 (AC002131) Strong similarity to extensin-like protein gb|Z34465 from Zea mays. [Arabidopsis thaliana] Length = 744
72	2028072	Tyr_Phospho_Site(809-817)
73	2028073	Pkc_Phospho_Site(15-17)
74	2028074	Pkc_Phospho_Site(13-15)
75	2028075	3E-14 >gb|AAD27733.1|AF132958_1 (AF132958) CGI-24 protein [Homo sapiens] Length = 241
76	2028076	Tyr_Phospho_Site(71-79)
77	2028077	3E-41 >emb|CAB10236.1| (Z97336) acylaminoacyl-peptidase like protein [Arabidopsis thaliana] Length = 426
78	2028078	Pkc_Phospho_Site(66-68)
79	2028079	7E-22 >ref|NP_009045.1|PTMF1| TATA element modulatory factor 1 >gi|423112|pir||A47212 transcription factor TMF, TATA element modulatory factor - human >gi|5870866|gb|AAD54608.1| (L01042) TATA element modulatory factor [Homo sapi
80	2028080	6E-24 >dbj|BAA25989| (D89051) ERD6 protein [Arabidopsis thaliana] Length = 496
81	2028081	Pkc_Phospho_Site(34-36)
82	2028082	Pkc_Phospho_Site(246-248)
83	2028083	1E-16 >gi|3883120 (AF082298) arabinogalactan-protein [Arabidopsis thaliana] Length = 131
84	2028084	1E-32 >gb|AAD26203.1|AF117267_1 (AF117267) UDP glucose:flavonoid 3-O-glucosyl transferase [Malus domestica] Length = 483
85	2028085	Tyr_Phospho_Site(497-504)
86	2028086	Pkc_Phospho_Site(393-395)
87	2028087	3' 8E-25 >gi|4093155 (AF088281) phytochrome-associated protein 1 [Arabidopsis thaliana] Length = 267
88	2028088	3' Pkc_Phospho_Site(18-20)
89	2028089	3' 1E-40 >gi|4006860|emb|CAB16778.1| (Z99707) thiol-disulfide interchange like protein [Arabidopsis thaliana] Length = 261
90	2028090	3' 5E-12 >gi|6225409|sp|O27955|GATA_ARCFU PROBABLE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT A (GLU-ADT SUBUNIT A) >gi|2648182 (AE000943) Glu-tRNA amidotransferase, subunit A (gatA-2) [Archaeoglobus fulgidus] Length = 457
91	2028091	3' 7E-57 >gi|4510424|gb|AAD21510.1| (AC006929) carboxypeptidase [Arabidopsis thaliana] Length = 361
92	2028092	3' Pkc_Phospho_Site(127-129)
93	2028093	5' 2E-70 >gi|1169598|sp|P46313|FD6E_ARATH OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM (DELTA-12 DESATURASE) >gi|438451 (L26296) delta-12 desaturase [Arabidopsis thaliana] Length = 383
94	2028094	5' Pkc_Phospho_Site(30-32)
95	2028095	5' Tyr_Phospho_Site(94-101)
96	2028096	5' Tyr_Phospho_Site(479-486)
97	2028097	5' 3E-17 >gi|6174930|sp|Q13200|PSD2_HUMAN 26S PROTEASOME REGULATORY SUBUNIT S2 (P97) (TUMOR NECROSIS FACTOR TYPE 1 RECEPTOR ASSOCIATED PROTEIN 2) (55.11 PROTEIN) Length = 908
98	2028098	5' Tyr_Phospho_Site(102-109)
99	2028099	5' 2E-28 >gi|2735764 (AF008651) MADS transcriptional factor; STMADS16 [Solanum tuberosum] Length = 234
100	2028100	5E-28 >gb|AAD43611.1|AC005698_10 (AC005698) T3P18.10 [Arabidopsis

		thaliana] Length = 482
101	2028101	Tyr_Phospho_Site(611-618)
102	2028102	9E-34 >gi|3687235 (AC005169) copia-like transposable element [Arabidopsis thaliana] Length = 213
103	2028103	3E-80 >emb|CAA76178.1| (Y16327) cyclic nucleotide-regulated ion channel [Arabidopsis thaliana] Length = 716
104	2028104	Tyr_Phospho_Site(1098-1104)
105	2028105	Tyr_Phospho_Site(164-172)
106	2028106	Pkc_Phospho_Site(15-17)
107	2028107	1E-126 >emb|CAA17550| (AL021961) receptor protein kinase - like protein [Arabidopsis thaliana] Length = 980
108	2028108	1E-51 >dbj|BAA77337.1| (AB019533) Nad-dependent formate dehydrogenase [Oryza sativa] Length = 376
109	2028109	2E-34 >emb|CAB16828.1| (Z99708) splicing factor-like protein [Arabidopsis thaliana] Length = 573
110	2028110	Tyr_Phospho_Site(69-76)
111	2028111	7E-54 >sp|Q96533|ADH3_ARATH GLUTATHIONE-DEPENDENT FORMALDEHYDE DEHYDROGENASE (FDH) (FALDH) (GSH-FDH) >gi|1498024 (U63931) glutathione-dependent formaldehyde dehydrogenase [Arabidopsis thaliana] Length = 379
112	2028112	2E-77) >emb|CAB10698| (Z97558) argininosuccinate lyase [Arabidopsis thaliana] Length = 517
113	2028113	Tyr_Phospho_Site(375-382)
114	2028114	3E-29 >pir||A42150 P-glycoprotein atpgp1 - Arabidopsis thaliana >gi|3849833|emb|CAA43646| (X61370) P-glycoprotein [Arabidopsis thaliana] >gi|4883607|gb|AAD31576.1|AC006922_8 (AC006922) P-glycoprotein pgp1 [Arabidopsis thaliana] Length = 1286
115	2028115	5E-46 >emb|CAB56614.1| (AJ234901) acetolactate synthase small subunit [Nicotiana plumbaginifolia] Length = 449
116	2028116	3' Pkc_Phospho_Site(24-26)
117	2028117	3' 2E-18 >gi|3249071 (AC004473) Contains similarity to protein-tyrosine phosphatase 2 gb|L15420 from Dictyostelium discoideum. EST gb|N38718 comes from this g [Arabidopsis thaliana] Length = 547
118	2028118	3' Tyr_Phospho_Site(25-33)
119	2028119	3' 1E-26 >gi|4531441|gb|AAD22126.1|AC006224_8 (AC006224) pectinesterase [Arabidopsis thaliana] Length = 518
120	2028120	3' Pkc_Phospho_Site(61-63)
121	2028121	5' Tyr_Phospho_Site(26-34)
122	2028122	5' 2E-36 >gi|3021279|emb|CAA18474.1| (AL022347) serine/threonine kinase [Arabidopsis thaliana] Length = 581
123	2028123	5' 1E-41 >gi|5454072|ref|NP_006416.1|pSLU7| step II splicing factor SLU7 >gi|4249705|gb|AAD13774.1| (AF101074) step II splicing factor SLU7 [Homo sapiens] Length = 586
124	2028124	5' Tyr_Phospho_Site(469-477)
125	2028125	5' Pkc_Phospho_Site(103-105)
126	2028126	1E-69 >emb|CAA17559| (AL021961) glucosyltransferase -like protein [Arabidopsis thaliana] Length = 478
127	2028127	Pkc_Phospho_Site(1-3)
128	2028128	Tyr_Phospho_Site(959-965)
129	2028129	1E-50 >gi|1432083 (U60981) homolog to Skp1p, an evolutionarily conserved kinetochore protein in budding yeast [Arabidopsis thaliana] >gi|3068807 (AF059294) Skp1 homolog [Arabidopsis thaliana] >gi|3719209

		(U97020) UIP1 [Arabidopsis thaliana] Length = 160
130	2028130	Pkc_Phospho_Site(42-44)
131	2028131	3E-30 >gi|1732515 (U62744) myosin heavy chain-like protein [Arabidopsis thaliana] Length = 209
132	2028132	2E-77 >dbj|BAA76297.1| (AB013912) DNA helicase [Mus musculus] Length = 463
133	2028133	1E-118 >sp|P32826|CBPX_ARATH SERINE CARBOXYPEPTIDASE PRECURSOR >gi|166674 (M81130) carboxypeptidase Y-like protein [Arabidopsis thaliana] >gi|445120|prf||1908426A carboxypeptidase Y [Arabidopsis thaliana] Length = 539
134	2028134	Pkc_Phospho_Site(67-69)
135	2028135	Pkc_Phospho_Site(1-3)
136	2028136	7E-74 >pir||S37495 peroxidase (EC 1.11.1.7) - Arabidopsis thaliana >gi|405611|emb|CAA50677| (X71794) peroxidase [Arabidopsis thaliana] Length = 353
137	2028137	1E-17 >gi|497174 (U07631) beta-hexosaminidase [Mus musculus] >gi|497196 (U07721) beta-hexosaminidase alpha-subunit [Mus musculus] Length = 528
138	2028138	Tyr_Phospho_Site(722-729)
139	2028139	2E-36 >gb|AAD14456| (AC005275) component of cytochrome B6-F complex [Arabidopsis thaliana] >gi|5725450|emb|CAB52433.1| (AJ243702) rieske iron-sulfur protein precursor [Arabidopsis thaliana] Length = 229
140	2028140	Pkc_Phospho_Site(23-25)
141	2028141	Tyr_Phospho_Site(57-64)
142	2028142	3' 2E-32 >gi|2499498|sp|Q42962|PGKY_TOBAC PHOSPHOGLYCERATE KINASE, CYTOSOLIC >gi|1161602|emb|CAA88840| (Z48976) phosphoglycerate kinase (PGK) [Nicotiana tabacum] Length = 401
143	2028143	3' 5E-15 >gi|3184098|emb|CAA19311.1| (AL023777) coenzyme a synthetase [Schizosaccharomyces pombe] Length = 512
144	2028144	3' Pkc_Phospho_Site(62-64)
145	2028145	5' Pkc_Phospho_Site(26-28)
146	2028146	5' 3E-80 >gi|3415115 (AF081202) villin 2 [Arabidopsis thaliana] Length = 976
147	2028147	5' Tyr_Phospho_Site(658-666)
148	2028148	5' Tyr_Phospho_Site(700-707)
149	2028149	1E-32 >gi|3193316 (AF069299) contains similarity to nucleotide sugar epimerases [Arabidopsis thaliana] Length = 430
150	2028150	Tyr_Phospho_Site(304-310)
151	2028151	Tyr_Phospho_Site(764-772)
152	2028152	3E-44 >gb|AAD27568.1|AF114171_9 (AF114171) H beta 58 homolog [Sorghum bicolor] Length = 616
153	2028153	8E-65 >gi|3249095 (AC003114) Contains similarity to dihydrofolate reductase (dfr1) gb|L13703 from Schizosaccharomyces pombe. ESTs gb|N37567 and gb|T43002 come from this gene. [Arabidopsis thaliana] Length = 550
154	2028154	2E-78 >gi|2281085 (AC002333) CTR1 protein kinase isolog [Arabidopsis thaliana] Length = 282
155	2028155	2E-84 >emb|CAB43938.1| (AJ006349) endo-beta-1,4-glucanase [Fragaria x ananassa] Length = 620
156	2028156	Tyr_Phospho_Site(253-260)
157	2028157	Rgd(302-304)
158	2028158	Tyr_Phospho_Site(762-769)
159	2028159	8E-87 >gb|AAD21729.1| (AC006931) citrate synthase [Arabidopsis thaliana] Length = 509
160	2028160	Tyr_Phospho_Site(64-72)

161	2028161	3E-89 >sp|P42749|UBC5_ARATH UBIQUITIN-CONJUGATING ENZYME E2-21 KD 2 (UBIQUITIN-PROTEIN LIGASE 5) (UBIQUITIN CARRIER PROTEIN 5) Length = 185
162	2028162	8E-91 >emb|CAA18628.1| (AL022580) pectinacetylesterase protein [Arabidopsis thaliana] Length = 362
163	2028163	Receptor_Cytokines_1(74-87)
164	2028164	3' 6E-38 >gi|3193301 (AF069298) Arabidopsis chloroplast outer envelope 86-like protein T10P11.19 (GB: AC002330) [Arabidopsis thaliana] Length = 1503
165	2028165	3' Rgd(776-778)
166	2028166	3' 2E-13 >gi|4337011|gb|AAD18035.1| (AF119572) zinc-binding peroxisomal integral membrane protein [Arabidopsis thaliana] Length = 381
167	2028167	5' Tyr_Phospho_Site(568-575)
168	2028168	5' Pkc_Phospho_Site(100-102)
169	2028169	Pkc_Phospho_Site(15-17)
170	2028170	4E-19 >gb|AAD22663.1|AC006555_1 (AC006555) beta-1,3-glucanase [Arabidopsis thaliana] >gi|4662638|gb|AAD26909.1|AC007233_1 (AC007233) beta-1,3-glucanase [Arabidopsis thaliana] Length = 473
171	2028171	4E-86 >pir||S44261 SRG1 protein - Arabidopsis thaliana >gi|479047|emb|CAA55654| (X79052) SRG1 [Arabidopsis thaliana] >gi|5734767|gb|AAD50032.1|AC007651_27 (AC007651) SRG1 Protein [Arabidopsis thaliana] Length = 358
172	2028172	1E-29 >gb|AAD22656.1|AC007138_20 (AC007138) NifU-like metallocluster assembly factor [Arabidopsis thaliana] Length = 174
173	2028173	1E-91 >gi|2062158 (AC001645) jasmonate inducible protein isolog [Arabidopsis thaliana] Length = 300
174	2028174	1E-101 >gb|AAF00639.1|AC009540_16 (AC009540) methionine synthase [Arabidopsis thaliana] Length = 765
175	2028175	2E-55 >sp|O64765|UAP1_ARATH PROBABLE UDP-N-ACETYLGLUCOSAMINE PYROPHOSPHORYLASE >gi|3033397 (AC004238) unknown protein [Arabidopsis thaliana] Length = 502
176	2028176	2E-20 >gi|1762933 (U66263) tumor-related protein [Nicotiana tabacum] Length = 210
177	2028177	2E-33 >gb|AAD24645.1|AC006220_1 (AC006220) symbiosis-related protein [Arabidopsis thaliana] Length = 120
178	2028178	Tyr_Phospho_Site(600-606)
179	2028179	8E-18 >gi|1840425 (U36586) alcohol dehydrogenase [Vitis vinifera] Length = 380
180	2028180	Tyr_Phospho_Site(339-345)
181	2028181	3' Tyr_Phospho_Site(368-375)
182	2028182	5' 4E-68 >gi|3914002|sp|O64948|LON1_ARATH MITOCHONDRIAL LON PROTEASE HOMOLOG 1 PRECURSOR >gi|2935279 (AF033862) Lon protease [Arabidopsis thaliana] Length = 888
183	2028183	5' Pkc_Phospho_Site(43-45)
184	2028184	5' 7E-51 >gi|3859659|emb|CAA20566.1| (AL031394) potassium transporter AtKT5p (AtKT5) [Arabidopsis thaliana] Length = 846
185	2028185	5' Pkc_Phospho_Site(60-62)
186	2028186	5' Rgd(273-275)
187	2028187	Pkc_Phospho_Site(30-32)
188	2028188	Pkc_Phospho_Site(57-59)
189	2028189	4E-32 >gi|2275196 (AC002337) water stress-induced protein, WSI76 isolog [Arabidopsis thaliana] >gi|4630746|gb|AAD26596.1|AC007236_1 (AC007236) water stress-induced protein [Arabidopsis thaliana] Length = 344

190	2028190	2E-14 >gi|2342666 (AF014502) seed coat peroxidase precursor [Glycine max] Length = 352
191	2028191	Tyr_Phospho_Site(150-156)
192	2028192	6E-41 >sp|P25865|UBC1_ARATH UBIQUITIN-CONJUGATING ENZYME E2-17 KD 1 (UBIQUITIN-PROTEIN LIGASE 1) (UBIQUITIN CARRIER PROTEIN 1) >gi|1076424|pir||S43781 ubiquitin-conjugating enzyme UBC1 - Arabidopsis thaliana >gi|442594|pdb|1AA
193	2028193	9E-43 >emb|CAA67551| (X99097) peroxidase [Arabidopsis thaliana] Length = 328
194	2028194	1E-158 >gi|3249096 (AC003114) Match to mRNA for importin alpha-like protein 4 (impa4) gb|Y14616 from A. thaliana. ESTs gb|N96440, gb|N37503, gb|N37498 and gb|T42198 come from this gene. [Arabidopsis thaliana] Length =
195	2028195	Tyr_Phospho_Site(41-48)
196	2028196	Tyr_Phospho_Site(33-41)
197	2028197	5E-29 >gi|2924788 (AC002334) similar to disease resistance protein [Arabidopsis thaliana] Length = 191
198	2028198	2E-54 >sp|P42804|HMA1_ARATH GLUTAMYL-TRNA REDUCTASE 1 PRECURSOR (GLUTR) >gi|454359 (U03774) glutamyl-tRNA reductase [Arabidopsis thaliana] Length = 543
199	2028199	3' Pkc_Phospho_Site(163-165)
200	2028200	3' 8E-65 >gi|6094242|sp|O23264|SBP_ARATH SELENIUM-BINDING PROTEIN >gi|2244759|emb|CAB10182.1| (Z97335) selenium-binding protein like [Arabidopsis thaliana] Length = 490
201	2028201	3' Tyr_Phospho_Site(558-566)
202	2028202	3' 1E-56 >gi|1483150|dbj|BAA12349| (D84417) monodehydroascorbate reductase [Arabidopsis thaliana] Length = 533
203	2028203	5' Tyr_Phospho_Site(569-575)
204	2028204	5' 5E-43 >gi|5262222|emb|CAB45848.1| (AL080254) reticuline oxidase-like protein [Arabidopsis thaliana] Length = 532
205	2028205	5' 1E-59 >gi|4337011|gb|AAD18035.1| (AF119572) zinc-binding peroxisomal integral membrane protein [Arabidopsis thaliana] Length = 381
206	2028206	5' 3E-61 >gi|1169601|sp|P46312|FD6C_ARATH OMEGA-6 FATTY ACID DESATURASE, CHLOROPLAST PRECURSOR >gi|493068 (U09503) chloroplast omega-6 fatty acid desaturase [Arabidopsis thaliana] Length = 418
207	2028207	Pkc_Phospho_Site(63-65)
208	2028208	Tyr_Phospho_Site(733-741)
209	2028209	Tyr_Phospho_Site(648-656)
210	2028210	3E-74 >gi|2347098 (U76845) ubiquitin-specific protease [Arabidopsis thaliana] >gi|4490742|emb|CAB38904.1| (AL035708) ubiquitin-specific protease (AtUBP3) [Arabidopsis thaliana] Length = 371
211	2028211	1E-89 >sp|P47927|AP2_ARATH FLORAL HOMEOTIC PROTEIN APETALA2 >gi|533709 (U12546) APETALA2 protein [Arabidopsis thaliana] >gi|2464888|emb|CAB16765.1| (Z99707) APETALA2 protein [Arabidopsis thaliana] Length = 432
212	2028212	Tyr_Phospho_Site(593-601)
213	2028213	Tyr_Phospho_Site(621-628)
214	2028214	Pkc_Phospho_Site(19-21)
215	2028215	1E-59 >gi|2688830 (AF000952) sugar transporter [Prunus armeniaca] Length = 475
216	2028216	Tyr_Phospho_Site(521-528)
217	2028217	Tyr_Phospho_Site(1176-1183)
218	2028218	Tyr_Phospho_Site(718-725)

219	2028219	Pkc_Phospho_Site(147-149)
220	2028220	Tyr_Phospho_Site(214-222)
221	2028221	2E-22 >sp|P11832|NIA1_ARATH NITRATE REDUCTASE 1 (NR1) >gi|486751|pir||S35228 nitrate reductase (NADH) (EC 1.6.6.1) 1 - Arabidopsis thaliana >gi|22757|emb|CAA79494| (Z19050) nitrate reductase [Arabidopsis thaliana] >gi|448286|prf||1916406A nitrate reductase [Arabidopsis thaliana] Length = 917
222	2028222	Tyr_Phospho_Site(217-224)
223	2028223	Tyr_Phospho_Site(875-882)
224	2028224	1E-27 >sp|Q39963|ER1_HEVBR ETHYLENE-INDUCIBLE PROTEIN HEVER >gi|2129913|pir||S60047 ethylene-responsive protein 1 - Para rubber tree >gi|1209317 (M88254) ethylene-inducible protein [Hevea brasiliensis] Length = 309
225	2028225	3' Pkc_Phospho_Site(43-45)
226	2028226	5' Pkc_Phospho_Site(85-87)
227	2028227	5' Tyr_Phospho_Site(679-686)
228	2028228	5' Uch_2_1(102-117)
229	2028229	5' 7E-23 >gi|2224933 (AF004216) ethylene-insensitive3 [Arabidopsis thaliana] >gi|2224935 (AF004217) ethylene-insensitive3 [Arabidopsis thaliana] Length = 628
230	2028230	Tyr_Phospho_Site(98-106)
231	2028231	6E-26 >dbj|BAA82637.1| (D63136) Beta-tubulin [Zinnia elegans] Length = 448
232	2028232	Pkc_Phospho_Site(68-70)
233	2028233	Tyr_Phospho_Site(718-726)
234	2028234	8E-52 >emb|CAA05875| (AJ003119) protein phosphatase 2C [Arabidopsis thaliana] Length = 511
235	2028235	3E-67 >sp|P45951|ARP_ARATH APURINIC ENDONUCLEASE-REDOX PROTEIN (DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE) >gi|472869|emb|CAA54234| (X76912) ARP protein [Arabidopsis thaliana] Length = 527
236	2028236	5E-76 >gb|AAF00669.1|AC008153_21 (AC008153) unknown protein [Arabidopsis thaliana] Length = 797
237	2028237	Pkc_Phospho_Site(45-47)
238	2028238	1E-55 >emb|CAB38817.1| (AL035679) fructose-bisphosphate aldolase [Arabidopsis thaliana] Length = 343
239	2028239	3E-56 >gb|AAD28617.1|AF129087_1 (AF129087) mitogen-activated protein kinase homologue [Medicago sativa] Length = 608
240	2028240	Pkc_Phospho_Site(17-19)
241	2028241	4E-29 >gb|AAF00639.1|AC009540_16 (AC009540) methionine synthase [Arabidopsis thaliana] Length = 765
242	2028242	3' Pkc_Phospho_Site(23-25)
243	2028243	3' 5E-17 >gi|5929906|gb|AAD56636.1|AF162150_1 (AF162150) COP1-interacting protein CIP8 [Arabidopsis thaliana] Length = 334
244	2028244	3' Tyr_Phospho_Site(566-573)
245	2028245	3' 1E-49 >gi|3256068|emb|CAA74397| (Y14068) Heat Shock Factor 3 [Arabidopsis thaliana] Length = 520
246	2028246	5' Pkc_Phospho_Site(165-167)
247	2028247	5' 4E-26 >gi|123078|sp|P13723|HEXA_DICDI BETA-HEXOSAMINIDASE ALPHA CHAIN PRECURSOR (N-ACETYL-BETA-GLUCOSAMINIDASE) (BETA-N-ACETYLHEXOSAMINIDASE) >gi|84092|pir||A30766 beta-N-acetylhexosaminidase (EC 3.2.1.52) A precursor - slime mold (Dictyostelium

		discoideum) >gi|167841 (J04065) beta-N-acetyl
248	2028248	5' Rgd(146-148)
249	2028249	5' Pkc_Phospho_Site(61-63)
250	2028250	2E-37 >emb|CAA09371.1| (AJ010829) GRAB1 protein [Triticum sp.] Length = 287
251	2028251	3E-84 >sp|P46644|AAT3_ARATH ASPARTATE AMINOTRANSFERASE, CHLOROPLAST PRECURSOR (TRANSAMINASE A) >gi|693692 (U15034) aspartate aminotransferase [Arabidopsis thaliana] Length = 449
252	2028252	2E-17 >gb|AAD11583.1|AAD11583 (AF071527) hypothetical protein [Arabidopsis thaliana] >gi|4262169|gb|AAD14469| (AC005275) hypothetical protein [Arabidopsis thaliana] Length = 236
253	2028253	1E-58 >pir||S57478 small GTP-binding protein - garden pea >gi|871508|emb|CAA90082| (Z49902) small GTP-binding protein [Pisum sativum] Length = 215
254	2028254	2E-21 >emb|CAA16710.1| (AL021687) RNase L inhibitor-like protein [Arabidopsis thaliana] Length = 600
255	2028255	1E-35 >emb|CAA16929.1| (AL021768) resistance protein RPP5-like [Arabidopsis thaliana] Length = 1715
256	2028256	8E-11 >sp|P49208|RK1_PEA 50S RIBOSOMAL PROTEIN L1, CHLOROPLAST PRECURSOR >gi|577089|emb|CAA58020| (X82776) chloroplast ribosomal protein L1 [Pisum sativum] Length = 208
257	2028257	2E-39 >dbj|BAA25989| (D89051) ERD6 protein [Arabidopsis thaliana] Length = 496
258	2028258	Pkc_Phospho_Site(71-73)
259	2028259	7E-77 >gi|3152587 (AC002986) Similar to CREB-binding protein homolog gb|U88570 from D. melanogaster and contains similarity to callus-associated protein gb|U01961 from Nicotiana tabacum. EST gb|W43427 comes from this gene. [Arabidopsis thaliana] Length = 1516
260	2028260	3E-27 >gi|4038040 (AC005936) proteinase inhibitor II [Arabidopsis thaliana] Length = 77
261	2028261	5E-54 >sp|P25851|F16P_ARATH FRUCTOSE-1,6-BISPHOSPHATASE, CHLOROPLAST PRECURSOR (D-FRUCTOSE-1,6-BISPHOSPHATE 1-PHOSPHOHYDROLASE) (FBPASE) >gi|99693|pir||S16582 fructose-bisphosphatase (EC 3.1.3.11) precursor, chloroplast - Arabidopsis thaliana >gi|11242|emb|CAA41154| (X58148) fructose-bisphosphatase [Arabidopsis thaliana] Length = 417
262	2028262	3' Pkc_Phospho_Site(20-22)
263	2028263	3' Tyr_Phospho_Site(469-475)
264	2028264	3' 4E-60 >gi|6358806|gb|AAF07386.1|AC010675_9 (AC010675) peptide transporter [Arabidopsis thaliana] Length = 644
265	2028265	5' Tyr_Phospho_Site(290-297)
266	2028266	5' Tyr_Phospho_Site(359-367)
267	2028267	5' Tyr_Phospho_Site(357-365)
268	2028268	5' 1E-11 >gi|1651723|dbj|BAA16651| (D90899) phosphoglycerate mutase [Synechocystis sp.] Length = 349
269	2028269	1E-101 >emb|CAB52675.1| (AJ010971) glucose-6-phosphate 1-dehydrogenase [Arabidopsis thaliana] Length = 515
270	2028270	Tyr_Phospho_Site(275-283)
271	2028271	3E-20 >gi|3252979 (AF068920) Ras-binding protein SUR-8 [Homo sapiens] >gi|3293320 (AF054828) leucine-rich repeat protein SHOC-2 [Homo sapiens] Length = 582
272	2028272	Pkc_Phospho_Site(137-139)
273	2028273	3E-44 >dbj|BAA06311| (D30622) novel serine/threonine protein kinase

		[Arabidopsis thaliana] Length = 421
274	2028274	Rgd(348-350)
275	2028275	6E-39 >sp|P81291|LE22_METJA 3-ISOPROPYLMALATE DEHYDRATASE LARGE SUBUNIT (ISOPROPYLMALATE ISOMERASE) (ALPHA-IPM ISOMERASE) (IPMI) >gi|2127740|pir||C64362 aconitate hydratase (EC 4.2.1.3) - Methanococcus jannaschii >gi|1591201 (U67499) 3-isopropylmalate dehydratase (leuC) [Methanococcus jannaschii] Length = 424
276	2028276	5E-83 >gi|4106395 (AF073744) raffinose synthase [Cucumis sativus] Length = 784
277	2028277	Pkc_Phospho_Site(19-21)
278	2028278	Tyr_Phospho_Site(541-548)
279	2028279	5E-39 >pdb|1SOX|A Chain A, Sulfite Oxidase From Chicken Liver >gi|3212611|pdb|1SOX|B Chain B, Sulfite Oxidase From Chicken Liver Length = 466
280	2028280	6E-45 >pir||S20940 DNA-binding protein - Arabidopsis thaliana Length = 246
281	2028281	Tyr_Phospho_Site(452-460)
282	2028282	1E-29 >gi|473874 (U08285) a membrane-associated salt-inducible protein [Nicotiana tabacum] Length = 435
283	2028283	Tyr_Phospho_Site(278-285)
284	2028284	9E-97 >gi|1173624 (U34744) cytochrome P-450 [Phalaenopsis sp. 'hybrid SM9108'] Length = 426
285	2028285	8E-89 >gi|1935914 (U77347) lethal leaf-spot 1 homolog [Arabidopsis thaliana] Length = 539
286	2028286	3' 2E-25 >gi|2323344 (AF014806) alpha-glucosidase 1 [Arabidopsis thaliana] Length = 902
287	2028287	3' Pkc_Phospho_Site(97-99)
288	2028288	3' 3E-12 >gi|6320470|ref|NP_010550.1|AKR1| Ankyrin repeat-containing protein; Akr1p >gi|728821|sp|P39010|AKR1_YEAST ANKYRIN REPEAT-CONTAINING PROTEIN AKR1 >gi|626094|pir||S48521 AKR1 protein - yeast (Saccharomyces cerevisiae) >gi|466522 (L31407) ankyrin repeat-containing protein [Saccharomyces cerevisiae] >gi|1230637 (U51030) Akr1p: Ankyrin repeat-containing protein (Swiss Prot. accession number P39010). [Saccharomyces cerevisiae] >gi|1586336|prf||2203403A ankyrin repeat-containing protein [Saccharomyces cerevisiae] Length = 764
289	2028289	3' Pkc_Phospho_Site(40-42)
290	2028290	3' 4E-43 >gi|4726118|gb|AAD28318.1|AC006436_9 (AC006436) somatic embryogenesis receptor-like kinase [Arabidopsis thaliana] Length = 520
291	2028291	5' Pkc_Phospho_Site(42-44)
292	2028292	5' 3E-15 >gi|4539386|emb|CAB37452.1| (AL035526) extensin-like protein [Arabidopsis thaliana] Length = 839
293	2028293	5' Tyr_Phospho_Site(720-727)
294	2028294	5' 2E-55 >gi|2129597|pir||S71217 glutamate dehydrogenase 1 - Arabidopsis thaliana >gi|1098960 (U37771) glutamate dehydrogenase 1 [Arabidopsis thaliana] >gi|1293095 (U53527) glutamate dehydrogenase 1 [Arabidopsis thaliana] Length = 411
295	2028295	3E-30 >gb|AAD34702.1|AC006341_30 (AC006341) Similar to gb|D14414 Indole-3-acetic acid induced protein from Vigna radiata. ESTs gb|AA712892 and gb|Z17613 come from this gene. [Arabidopsis thaliana] Length = 147
296	2028296	3E-11 >pir||S47536 SWH1 protein (version 2) - yeast (Saccharomyces cerevisiae) >gi|402658|emb|CAA52646| (X74552) SWH1 [Saccharomyces

		cerevisiae] >gi|1090523|prf||2019253A oxysterol-binding protein-like protein [Saccharomyces cerevisiae] Length = 1190
297	2028297	Tyr_Phospho_Site(8-16)
298	2028298	Tyr_Phospho_Site(397-404)
299	2028299	Pkc_Phospho_Site(15-17)
300	2028300	Pkc_Phospho_Site(221-223)
301	2028301	6E-43 >pir||S71229 RNA-binding protein 37 - Arabidopsis thaliana >gi|1174153 (U44134) RNA-binding protein [Arabidopsis thaliana] Length = 336
302	2028302	Tyr_Phospho_Site(817-824)
303	2028303	1E-90 >emb|CAB43971.1| (AL078579) beta-glucosidase [Arabidopsis thaliana] Length = 517
304	2028304	Pkc_Phospho_Site(43-45)
305	2028305	Tyr_Phospho_Site(45-51)
306	2028306	3' Tyr_Phospho_Site(208-215)
307	2028307	3' 2E-33 >gi|3776572 (AC005388) ESTs gb|R65052, gb|AA712146, gb|H76533, gb|H76282, gb|AA650771, gb|H76287, gb|AA650887, gb|N37383, gb|Z29721 and gb|Z29722 come from this gene. [Arabidopsis thaliana] Length = 285
308	2028308	3' 7E-11 >gi|3560235|emb|CAA20703.1| (AL031530) hypothetical zinc finger protein [Schizosaccharomyces pombe] Length = 680
309	2028309	5' Pkc_Phospho_Site(39-41)
310	2028310	5' Tyr_Phospho_Site(310-317)
311	2028311	5' Pkc_Phospho_Site(84-86)
312	2028312	5' Pkc_Phospho_Site(16-18)
313	2028313	Pkc_Phospho_Site(20-22)
314	2028314	1E-12 >gi|154692 (M73322) cellulase E-4 [Thermomonospora fusca] Length = 376
315	2028315	Pkc_Phospho_Site(92-94)
316	2028316	2E-60 >gi|2462824 (AF000657) similar to Jun activation domain binding protein [Arabidopsis thaliana] >gi|2791885 (AF042334) JAB1 [Arabidopsis thaliana] Length = 357
317	2028317	Tyr_Phospho_Site(725-733)
318	2028318	4E-54) >gb|AAD48837.1|AF166351_1 (AF166351) alanine:glyoxylate aminotransferase 2 homolog [Arabidopsis thaliana] Length = 476
319	2028319	6E-43 >sp|P42731|PAB2_ARATH POLYADENYLATE-BINDING PROTEIN 2 (POLY(A) BINDING PROTEIN 2) (PABP 2) >gi|304109 (L19418) poly(A)-binding protein [Arabidopsis thaliana] >gi|2911051|emb|CAA17561| (AL021961) poly(A)-binding protein [
320	2028320	Pkc_Phospho_Site(41-43)
321	2028321	2E-20 >dbj|BAA25989| (D89051) ERD6 protein [Arabidopsis thaliana] Length = 496
322	2028322	6E-43 >sp|Q42208|RL7_ARATH 60S RIBOSOMAL PROTEIN L7 >gi|3212879 (AC004005) ribosomal protein L7 [Arabidopsis thaliana] Length = 247
323	2028323	4E-11 >emb|CAB53646.1| (AL110123) multidrug resistance protein/P-glycoprotein-like [Arabidopsis thaliana] Length = 1222
324	2028324	3' 4E-18 >gi|3941528 (AF062918) transcription factor [Arabidopsis thaliana] Length = 335
325	2028325	3' Tyr_Phospho_Site(808-815)
326	2028326	3' 1E-19 >gi|1694711|emb|CAA70769| (Y09581) FRO1 [Arabidopsis thaliana] Length = 704
327	2028327	3' 8E-12 >gi|2894597|emb|CAA17131.1| (AL021889) bHLH protein-like [Arabidopsis thaliana] Length = 589

328	2028328	3' 3E-28 >gi|461812|sp|Q05047|CP72_CATRO CYTOCHROME P450 72A1 (CYPLXXII) (PROBABLE GERANIOL-10-HYDROXYLASE) (GE10H) >gi|167484 (L10081) Cytochrome P-450 protein [Catharanthus roseus] >gi|445604|prf||1909351A cytochrome P450 [Catharanthus roseus] Length = 524
329	2028329	3' 2E-15 >gi|400972|sp|P30986|RETO_ESCCA RETICULINE OXIDASE PRECURSOR (BERBERINE-BRIDGE-FORMING ENZYME) (BBE) (TETRAHYDROPROTOBERBERINE SYNTHASE) >gi|99506|pir||A41533 reticuline oxidase (EC 1.5.3.9) precursor - California poppy >gi|239110|bbs|65555 (S65550) (S)-reticuline:oxygen oxidoreductas
330	2028330	5' Tyr_Phospho_Site(71-79)
331	2028331	5' 2E-69 >gi|123340|sp|P14891|HMD1_ARATH 3-HYDROXY-3-METHYLGLUTARYL-COENZYME A REDUCTASE 1 (HMG-COA REDUCTASE 1) (HMGR1) >gi|99714|pir||A32107 hydroxymethylglutaryl-CoA reductase (NADPH) (EC 1.1.1.34) - Arabidopsis thaliana >gi|16336|emb|CAA33139| (X15032) hydroxy methylglutaryl CoA reductase
332	2028332	5' 3E-19 >gi|5731257|gb|AAD48836.1|AF165924_1 (AF165924) auxin-induced basic helix-loop-helix transcription factor [Gossypium hirsutum] Length = 314
333	2028333	5' Pkc_Phospho_Site(22-24)
334	2028334	5' Tyr_Phospho_Site(17-24)
335	2028335	Tyr_Phospho_Site(1196-1204)
336	2028336	1E-53 >sp|Q08467|KC21_ARATH CASEIN KINASE II, ALPHA CHAIN 1 (CK II) >gi|419752|pir||S31098 casein kinase II (EC 2.7.1.-) alpha-type chain (clone ATCKA1) - Arabidopsis thaliana >gi|391603|dbj|BAA01090| (D10246) casein kinase II catalytic subunit [Arabidopsis thaliana] Length = 33
337	2028337	4E-28 >sp|O24164|PPOM_TOBAC PROTOPORPHYRINOGEN OXIDASE, MITOCHONDRIAL (PPO II) (PROTOPORPHYRINOGEN IX OXIDASE ISOZYME II) (PPX II) >gi|2370335|emb|CAA73866| (Y13466) protoporphyrinogen oxidase [Nicotiana tabacum] >gi|3929920|dbj|BAA34712| (AB020500) mitochondrial protoporphyrino
338	2028338	2E-50 >emb|CAA63010| (X91917) LEA D113 homologue type2 [Arabidopsis thaliana] >gi|3668076 (AC004667) LEA D113 type2 protein [Arabidopsis thaliana] Length = 97
339	2028339	4E-12 >gi|2224915 (U95968) beta-expansin [Oryza sativa] Length = 261
340	2028340	Tyr_Phospho_Site(494-501)
341	2028341	3E-18 >sp|P54926|MYO1_LYCES MYO-INOSITOL-1(OR 4)-MONOPHOSPHATASE 1 (IMP 1) (INOSITOL MONOPHOSPHATASE 1) >gi|1098977 (U39444) myo-inositol monophosphatase 1 [Lycopersicon esculentum] Length = 273
342	2028342	Pkc_Phospho_Site(8-10)
343	2028343	Pkc_Phospho_Site(13-15)
344	2028344	Tyr_Phospho_Site(665-672)
345	2028345	9E-65 >emb|CAA74028.1| (Y13694) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana] >gi|2827525|emb|CAA16533.1| (AL021633) multicatalytic endopeptidase complex, proteasome precu
346	2028346	Tyr_Phospho_Site(314-321)
347	2028347	1E-52 >gi|3540183 (AC004122) Highly Similar to branched-chain amino acid aminotransferase [Arabidopsis thaliana] Length = 318
348	2028348	2E-11 >emb|CAB10522.1| (Z97343) DNA-binding protein homolog [Arabidopsis thaliana] Length = 459
349	2028349	7E-15 >emb|CAA69072| (Y07765) S-adenosylmethionine decarboxylase [Arabidopsis thaliana] Length = 51
350	2028350	2E-27 >sp|P19954|RR30_SPIOL 30S RIBOSOMAL PROTEIN S30,

		CHLOROPLAST PRECURSOR (CS-S5) (CS5) (S22) (RIBOSOMAL PROTEIN 1) (PSRP-1) >gi|279640|pir||R3SPS5 ribosomal protein CS-S22 precursor, chloroplast - spinach >gi|12316|emb|CAA41960| (X59270) chloroplast ribosomal protein S22 [Spinacia oleracea] >gi|18031|emb|CAA33403| (X15344) spinach S22 r-protein [Spinacia oleracea] Length = 302
351	2028351	3' Tyr_Phospho_Site(344-350)
352	2028352	5' 3E-65 >gi|3164126|dbj|BAA28531| (D78598) cytochrome P450 monooxygenase [Arabidopsis thaliana] >gi|5262761|emb|CAB45909.1| (AL080283) cytochrome P450 monooxygenase [Arabidopsis thaliana] Length = 499
353	2028353	5' 1E-76 >gi|5915830|sp|Q96514|C7B7_ARATH CYTOCHROME P450 71B7 >gi|1523796|emb|CAA66458| (X97864) cytochrome P450 [Arabidopsis thaliana] >gi|4850394|gb|AAD31064.1|AC007357_13 (AC007357) Identical to gb|X97864 cytochrome P450 from Arabidopsis thaliana and is a member of the PF|00067 Cytochrome
354	2028354	5' Tyr_Phospho_Site(209-216)
355	2028355	5' Tyr_Phospho_Site(823-831)
356	2028356	5' Pkc_Phospho_Site(6-8)
357	2028357	5' 3E-45 >gi|5541691|emb|CAB51197.1| (AL096859) glucuronosyl transferase-like protein (fragment) [Arabidopsis thaliana] Length = 271
358	2028358	4E-39 >gi|3201623 (AC004669) shaggy-like kinase dzeta [Arabidopsis thaliana] Length = 412
359	2028359	Pkc_Phospho_Site(2-4)
360	2028360	Tyr_Phospho_Site(638-645)
361	2028361	Tyr_Phospho_Site(297-304)
362	2028362	5E-83 >gi|2275196 (AC002337) water stress-induced protein, WSI76 isolog [Arabidopsis thaliana] >gi|4630746|gb|AAD26596.1|AC007236_1 (AC007236) water stress-induced protein [Arabidopsis thaliana] Length = 344
363	2028363	4E-76) >gb|AAD20113| (AC006304) proline iminopeptidase [Arabidopsis thaliana] Length = 329
364	2028364	1E-48 >emb|CAA66964| (X98320) peroxidase [Arabidopsis thaliana] >gi|1429215|emb|CAA67310| (X98774) peroxidase ATP6a [Arabidopsis thaliana] Length = 336
365	2028365	3E-31 >gb|AAB95298.1| (AC003105) beta-ketoacyl-CoA synthase [Arabidopsis thaliana] Length = 509
366	2028366	Tyr_Phospho_Site(370-378)
367	2028367	1E-39 >emb|CAA65384| (X96539) malate dehydrogenase [Mesembryanthemum crystallinum] Length = 332
368	2028368	3' Tyr_Phospho_Site(176-183)
369	2028369	3' Pkc_Phospho_Site(10-12)
370	2028370	3' 2E-52 >gi|2739376 (AC002505) permease [Arabidopsis thaliana] Length = 551
371	2028371	3' 2E-53 >gi|2316016 (U92650) MRP-like ABC transporter [Arabidopsis thaliana] Length = 1515
372	2028372	3' Tyr_Phospho_Site(414-420)
373	2028373	5' Tyr_Phospho_Site(10-17)
374	2028374	5' 5E-77 >gi|2129553|pir||S71774 calcium-dependent protein kinase 6 - Arabidopsis thaliana Length = 529
375	2028375	5' Pkc_Phospho_Site(53-55)
376	2028376	5' 1E-42 >gi|1495768|emb|CAA92823| (Z68506) chloroplast inner envelope protein, 110 kD (IEP110) [Pisum sativum] Length = 996
377	2028377	5' 2E-75 >gi|3914425|sp|O23717|PRCE_ARATH PROTEASOME EPSILON CHAIN PRECURSOR (MACROPAIN EPSILON CHAIN) (MULTICATALYTIC ENDOPEPTIDASE COMPLEX EPSILON CHAIN) >gi|2511596|emb|CAA74029.1|

		(Y13695) multicatalytic endopeptidase complex, proteasome precursor, beta subunit [Arabidopsis thaliana] >gi|
378	2028378	3E-48) >gi|2088650 (AF002109) peroxisomal ATP/ADP carrier protein isolog [Arabidopsis thaliana] Length = 331
379	2028379	Pkc_Phospho_Site(40-42)
380	2028380	3E-16 >gb|AAD39612.1|AC007454_11 (AC007454) Similar to gb|X92204 NAM gene product from Petunia hybrida. ESTs gb|H36656 and gb|AA651216 come from this gene. [Arabidopsis thaliana] Length = 557
381	2028381	8E-79 >emb|CAA65051| (X95736) amino acid permease 6 [Arabidopsis thaliana] Length = 481
382	2028382	Pkc_Phospho_Site(65-67)
383	2028383	3E-18 >gb|AAD46412.1|AF096262_1 (AF096262) ER6 protein [Lycopersicon esculentum] Length = 168
384	2028384	1E-81 >gi|2827139 (AF027172) cellulose synthase catalytic subunit [Arabidopsis thaliana] >gi|4049343|emb|CAA22568.1| (AL034567) cellulose synthase catalytic subunit (RSW1) [Arabidopsis thaliana] Length = 1081
385	2028385	Pkc_Phospho_Site(9-11)
386	2028386	6E-13 >gi|2342674 (AC000106) Similar to ATP-dependent Clp protease (gb|D90915). EST gb|N65461 comes from this gene. [Arabidopsis thaliana] Length = 292
387	2028387	7E-46 >gb|AAD29776.1|AF074021_8 (AF074021) symbiosis-related protein [Arabidopsis thaliana] Length = 122
388	2028388	4E-41 >dbj|BAA07555| (D38552) The ha1539 protein is related to cyclophilin. [Homo sapiens] Length = 645
389	2028389	Tyr_Phospho_Site(858-864)
390	2028390	1E-49 >pir||S71265 ferritin - Arabidopsis thaliana >gi|1246401|emb|CAA63932| (X94248) ferritin [Arabidopsis thaliana] Length = 255
391	2028391	Tyr_Phospho_Site(582-588)
392	2028392	3' Pkc_Phospho_Site(34-36)
393	2028393	3' Tyr_Phospho_Site(231-239)
394	2028394	3' Pkc_Phospho_Site(31-33)
395	2028395	3' 6E-25 >gi|2098713 (U82977) pectinesterase [Citrus sinensis] Length = 510
396	2028396	3' Tyr_Phospho_Site(93-100)
397	2028397	5' Tyr_Phospho_Site(287-293)
398	2028398	5' Pkc_Phospho_Site(22-24)
399	2028399	5' Pkc_Phospho_Site(37-39)
400	2028400	5' 2E-36 >gi|1170170|sp|P46602|HAT3_ARATH HOMEOBOX-LEUCINE ZIPPER PROTEIN HAT3 (HD-ZIP PROTEIN 3) >gi|549889 (U09338) homeobox protein [Arabidopsis thaliana] >gi|549890 (U09339) homeobox protein [Arabidopsis thaliana] Length = 315
401	2028401	Tyr_Phospho_Site(384-390)
402	2028402	1E-54 >sp|P43188|KADC_MAIZE ADENYLATE KINASE, CHLOROPLAST (ATP-AMP TRANSPHOSPHORYLASE) >gi|629863|pir||S45634 adenylate kinase (EC 2.7.4.3), chloroplast - maize >gi|3114421|pdb|1ZAK|A Chain A, Adenylate Kinase From Maize In Complex With The Inhibitor P1,P5-Bis(Adenosine-5'-)pentaphosphate (Ap5a) >gi|3114422|pdb|1ZAK|B Chain B, Adenylate Kinase From Maize In Complex With The Inhibitor P1,P5-Bis(Adenosine-5'-)pentaphosphate (Ap5a) Length = 222
403	2028403	1E-101 >sp|P54888|P5C2_ARATH DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHETASE B (P5CS B) [INCLUDES: GLUTAMATE 5-KINASE (GAMMA-GLUTAMYL KINASE) (GK); GAMMA-GLUTAMYL PHOSPHATE REDUCTASE

		(GPR) (GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE) (GLUTAMYL-GAMMA-SEMIALDE... >gi|887388|emb|CAA60447| (X86778) pyrroline-5-carboxylate synthetase B [Arabidopsis thaliana] >gi|1669658|emb|CAA70527| (Y09355) pyrroline-5-carboxlyate synthetase [Arabidopsis thaliana] Length = 726
404	2028404	Tyr_Phospho_Site(585-592)
405	2028405	6E-40 >pir||HSWT4 histone H4 - wheat >gi|70773|pir||HSPM4 histone H4 - garden pea Length = 102
406	2028406	Tyr_Phospho_Site(329-336)
407	2028407	Pkc_Phospho_Site(117-119)
408	2028408	3E-93 >gb|AAD16946| (AF106324) sodium proton exchanger Nhx1 [Arabidopsis thaliana] Length = 538
409	2028409	Tyr_Phospho_Site(852-860)
410	2028410	Pkc_Phospho_Site(66-68)
411	2028411	3' 2E-18 >gi|629728|pir||S46959 porin I, 36K - potato >gi|1076680|pir||C55364 porin (clone pPOM 36.1) - potato mitochondrion >gi|515358|emb|CAA56601| (X80388) 36kDa porin I [Solanum tuberosum] Length = 276
412	2028412	3' Tyr_Phospho_Site(330-337)
413	2028413	3' Tyr_Phospho_Site(208-215)
414	2028414	3' Pkc_Phospho_Site(55-57)
415	2028415	3' 1E-23 >gi|2499535|sp|Q41364|SOT1_SPIOL 2-OXOGLUTARATE/MALATE TRANSLOCATOR PRECURSOR >gi|595681 (U13238) 2-oxoglutarate/malate translocator [Spinacia oleracea] Length = 569
416	2028416	3' 1E-10 >gi|99749|pir||S20918 probable serine/threonine-specific protein kinase ATPK64 (EC 2.7.1.-) - Arabidopsis thaliana >gi|217843|dbj|BAA01731| (D10937) protein kinase [Arabidopsis thaliana] Length = 498
417	2028417	3' Tyr_Phospho_Site(693-701)
418	2028418	3' Pkc_Phospho_Site(115-117)
419	2028419	5' Pkc_Phospho_Site(2-4)
420	2028420	5' 6E-77 >gi|5730139|emb|CAB52472.1| (AJ243705) ferredoxin-NADP+ reductase [Arabidopsis thaliana] Length = 360
421	2028421	5' Rgd(605-607)
422	2028422	8E-13 >gb|AAD41415.1|AC007727_4 (AC007727) Contains similarity to gb|U07707 epidermal growth factor receptor substrate (eps15) from Homo sapiens and contains 2 PF|00036 EF hand domains. ESTs gb|T44428 and gb|AA395440 come from this gene. [Arabidop... Length = 1181
423	2028423	Tyr_Phospho_Site(412-419)
424	2028424	9E-72 >gb|AAD32285.1|AC006533_9 (AC006533) poly(ADP-ribose) glycohydrolase [Arabidopsis thaliana] Length = 997
425	2028425	Tyr_Phospho_Site(77-84)
426	2028426	Tyr_Phospho_Site(800-807)
427	2028427	1E-22 >ref|NP_004658.1|PHERC2| hect domain and RLD 2 >gi|4079809|gb|AAD08657.1| (AF071172) HERC2 [Homo sapiens] Length = 4834
428	2028428	Rgd(235-237)
429	2028429	Tyr_Phospho_Site(399-406)
430	2028430	6E-54 >gi|2739368 (AC002505) cyclin-like protein [Arabidopsis thaliana] Length = 361
431	2028431	6E-46 >gb|AAD21729.1| (AC006931) citrate synthase [Arabidopsis thaliana] Length = 509
432	2028432	2E-45 >gi|2459448 (AC002332) cinnamoyl-CoA reductase [Arabidopsis thaliana] Length = 321
433	2028433	1E-27 >gb|AAD39990.1|AF150083_1 (AF150083) small zinc finger-like protein [Arabidopsis thaliana] Length = 77

434	2028434	2E-44 >gi|2829133 (AF043351) adenosine-5'-phosphosulfate-kinase [Arabidopsis thaliana] >gi|4490745|emb|CAB38907.1| (AL035708) adenosine-5'-phosphosulfate-kinase [Arabidopsis thaliana] Length = 293
435	2028435	Pkc_Phospho_Site(21-23)
436	2028436	2E-47 >dbj|BAA77358.1| (AB020023) DNA-binding protein NtWRKY3 [Nicotiana tabacum] Length = 328
437	2028437	Pkc_Phospho_Site(41-43)
438	2028438	3' Tyr_Phospho_Site(28-35)
439	2028439	3' Tyr_Phospho_Site(210-217)
440	2028440	3' 5E-18 >gi|2827665|emb|CAA16619.1| (AL021637) vacuolar sorting receptor-like protein [Arabidopsis thaliana] Length = 626
441	2028441	3' 3E-25 >gi|1419090|emb|CAA64422| (X94968) 37kDa chloroplast inner envelope membrane polypeptide precursor [Nicotiana tabacum] Length = 335
442	2028442	3' Tyr_Phospho_Site(681-688)
443	2028443	3' 8E-69 >gi|5921663|gb|AAD56290.1|AF162279_1 (AF162279) 10-formyltetrahydrofolate synthetase [Arabidopsis thaliana] Length = 634
444	2028444	5' Tyr_Phospho_Site(422-428)
445	2028445	5' 7E-53 >gi|3914097|sp|O49071|MYOP_MESCR MYO-INOSITOL-1(OR 4)-MONOPHOSPHATASE (IMP) (INOSITOL MONOPHOSPHATASE) >gi|2708322 (AF037220) inositol monophosphatase [Mesembryanthemum crystallinum] Length = 270
446	2028446	5' 2E-26 >gi|2921323|gb|AAC04713.1| (AF034112) beta-1,3-glucanase 7 [Glycine max] Length = 245
447	2028447	5' Tyr_Phospho_Site(102-109)
448	2028448	Pkc_Phospho_Site(17-19)
449	2028449	Tyr_Phospho_Site(658-664)
450	2028450	1E-23 >emb|CAA18991| (AL023518) transport protein [Schizosaccharomyces pombe] Length = 397
451	2028451	1E-106 >gi|2737926 (U77673) fimbrin-like protein AtFim2 [Arabidopsis thaliana] Length = 456
452	2028452	4E-84 >gi|3643604 (AC005395) receptor-like protein kinase [Arabidopsis thaliana] Length = 960
453	2028453	Pkc_Phospho_Site(7-9)
454	2028454	Tyr_Phospho_Site(1156-1162)
455	2028455	2E-70 >gi|4098521 (U79160) HMG-CoA synthase [Arabidopsis thaliana] >gi|4098523 (U79161) HMG-CoA synthase [Arabidopsis thaliana] >gi|5002517|emb|CAB44320.1| (AL078606) hydroxymethylglutaryl-CoA synthase [Arabidopsis thaliana] Length = 461
456	2028456	3E-72 >gi|2583111 (AC002387) dihydrodipicolinate synthase [Arabidopsis thaliana] Length = 365
457	2028457	9E-79) >emb|CAA35887| (X51514) precursor acetolactate synthase (670 AA) [Arabidopsis thaliana] Length = 670
458	2028458	4E-86) >dbj|BAA84380.1| (AP000423) PSII D2 protein [Arabidopsis thaliana] Length = 353
459	2028459	2E-67 >emb|CAA76758.1| (Y17386) In2.1 protein [Triticum aestivum] Length = 243
460	2028460	Pkc_Phospho_Site(45-47)
461	2028461	Tyr_Phospho_Site(349-357)
462	2028462	Tyr_Phospho_Site(303-310)
463	2028463	3' 1E-28 >gi|4106340|gb|AAD02810| (AF062396) protein phosphatase 2A regulatory subunit isoform B' delta [Arabidopsis thaliana] Length = 477
464	2028464	3' 5E-41 >gi|4185133 (AC005724) zinc finger protein [Arabidopsis thaliana] Length = 181

465	2028465	3' 1E-44 >gi|4678357|emb|CAB41167.1| (AL049659) cytochrome P450-like protein [Arabidopsis thaliana] Length = 490
466	2028466	5' Pkc_Phospho_Site(82-84)
467	2028467	5' 1E-31 >gi|2500185|sp|Q23862|RACE_DICDI RAS-RELATED PROTEIN RACE >gi|1373067 (U41222) RacE [Dictyostelium discoideum] Length = 223
468	2028468	5' 8E-74 >gi|4587685|gb|AAD25855.1|AC007197_8 (AC007197) methylmalonate semi-aldehyde dehydrogenase [Arabidopsis thaliana] Length = 607
469	2028469	5' 2E-72 >gi|2494174|sp|Q42521|DCE1_ARATH GLUTAMATE DECARBOXYLASE 1 (GAD 1) >gi|497979 (U10034) glutamate decarboxylase [Arabidopsis thaliana] Length = 502
470	2028470	5' 6E-75 >gi|5669047|gb|AAD46145.1| (AF081573) 19S proteasome regulatory complex subunit S6A [Arabidopsis thaliana] Length = 424
471	2028471	5' Pkc_Phospho_Site(20-22)
472	2028472	5' 3E-71 >gi|2501056|sp|Q39230|SYS_ARATH SERYL-TRNA SYNTHETASE (SERINE—TRNA LIGASE) (SERRS) >gi|2129737|pir||S71293 seryl-tRNA synthetase - Arabidopsis thaliana >gi|1359497|emb|CAA94388| (Z70313) seryl-tRNA Synthetase [Arabidopsis thaliana] Length = 451
473	2028473	Pkc_Phospho_Site(49-51)
474	2028474	Pkc_Phospho_Site(26-28)
475	2028475	Tyr_Phospho_Site(217-225)
476	2028476	4E-81 >emb|CAA67336| (X98804) peroxidase ATP18a [Arabidopsis thaliana] Length = 346
477	2028477	4E-34 >sp|P56286|IF2A_SCHPO EUKARYOTIC TRANSLATION INITIATION FACTOR 2 ALPHA SUBUNIT (EIF-2-ALPHA) >gi|2706460|emb|CAA15918.1| (AL021046) eukaryotic translation initiation factor 2 alpha subunit [Schizosaccharomyces pombe] Length = 306
478	2028478	1E-117 >sp|P54609|CC48_ARATH CELL DIVISION CYCLE PROTEIN 48 HOMOLOG >gi|2118115|pir||S60112 cell division control protein CDC48 homolog - Arabidopsis thaliana >gi|1019904 (U37587) cell division cycle protein [Arabidopsis thaliana] Length = 809
479	2028479	2E-84 >emb|CAA23006| (AL035356) mitochondrial uncoupling protein [Arabidopsis thaliana] Length = 313
480	2028480	7E-11 >gi|3335347 (AC004512) Contains similarity to ARI, RING finger protein gb|X98309 from Drosophila melanogaster. ESTs gb|T44383, gb|W43120, gb|N65868, gb|H36013, gb|AA042241, gb|T76869 and gb|AA042359 come from this gene. [Arabidopsis thaliana] Length = 644
481	2028481	1E-63 >gi|682728 (L40031) S-adenosyl-L-methionine:trans-caffeoyl-Coenzyme A 3-O-methyltransferase [Arabidopsis thaliana] Length = 212
482	2028482	1E-22 >gi|3687243 (AC005169) ribosomal protein [Arabidopsis thaliana] Length = 68
483	2028483	7E-42 >gi|3415115 (AF081202) villin 2 [Arabidopsis thaliana] Length = 976
484	2028484	Tyr_Phospho_Site(204-211)
485	2028485	Tyr_Phospho_Site(58-65)
486	2028486	3' 8E-18 >gi|2804278|dbj|BAA24448| (AB003516) squalene epoxidase [Panax ginseng] Length = 539
487	2028487	3' 5E-20 >gi|3914394|sp|Q42908|PMGI_MESCR 2,3-BISPHOSPHOGLYCERATE-INDEPENDENT PHOSPHOGLYCERATE MUTASE (PHOSPHOGLYCEROMUTASE) (BPG-INDEPENDENT PGAM) (PGAM-I) >gi|2118335|pir||S60473 phosphoglycerate mutase (EC 5.4.2.1) - common ice plant >gi|602426 (U16021) phosphoglyceromutase [Mesembryanthemum crystallinum] Length = 559
488	2028488	3' Wd_Repeats(594-608)

489	2028489	3' Pkc_Phospho_Site(4-6)
490	2028490	5' 6E-69 >gi|5738864|emb|CAA63220.1| (X92486) isocitrate dehydrogenase (NAD+) [Solanum tuberosum] Length = 470
491	2028491	5' 2E-74 >gi|4927412|gb|AAD33097.1|AF082525_1 (AF082525) homoserine kinase [Arabidopsis thaliana] Length = 370
492	2028492	5' 1E-60 >gi|3128168 (AC004521) carboxyl-terminal peptidase [Arabidopsis thaliana] Length = 415
493	2028493	5' Pkc_Phospho_Site(41-43)
494	2028494	5' 3E-62 >gi|4006869|emb|CAB16787.1| (Z99707) patatin-like protein [Arabidopsis thaliana] Length = 414
495	2028495	3E-18 >gi|3139079 (AF062537) cullin 3 [Homo sapiens] Length = 768
496	2028496	Tyr_Phospho_Site(1069-1076)
497	2028497	1E-63 >gb|AAC27707.1| (AF067789) tSNARE AtTLG2a [Arabidopsis thaliana] Length = 322
498	2028498	9E-36 >gi|4091806 (AF052585) CONSTANS-like protein 2 [Malus domestica] Length = 329
499	2028499	9E-21 >gi|2191133 (AF007269) Arabidopsis thaliana G-box binding factor 2 (SP:P42774) [Arabidopsis thaliana] Length = 380
500	2028500	4E-50 >gi|3650032 (AC005396) gibberellin-regulated protein GAST1-like [Arabidopsis thaliana] Length = 108
501	2028501	1E-27 >sp|Q96330|FLAV_ARATH FLAVONOL SYNTHASE (FLS) >gi|1628622 (U72631) flavonol synthase [Arabidopsis thaliana] >gi|1805305 (U84258) flavonol synthase [Arabidopsis thaliana] >gi|1805307 (U84259) flavonol synthase [Arabidopsis thaliana] >gi|1805309 (U84260) flavonol synthase [Arabidopsis thaliana] Length = 336
502	2028502	4E-61 >gi|3176686 (AC003671) Similar to high affinity potassium transporter, HAK1 protein gb|U22945 from Schwanniomyces occidentalis. [Arabidopsis thaliana] Length = 764
503	2028503	4E-61 >sp|P15455|12S1_ARATH 12S SEED STORAGE PROTEIN PRECURSOR >gi|81604|pir||S08509 cruciferin precursor (CRA1) - Arabidopsis thaliana >gi|166676 (M37247) 12S storage protein CRA1 [Arabidopsis thaliana] >gi|808936|emb|CAA3249
504	2028504	Tyr_Phospho_Site(13-20)
505	2028505	3E-39 >gi|2062164 (AC001645) jasmonate inducible protein isolog [Arabidopsis thaliana] Length = 470
506	2028506	1E-82) >sp|P32962|NRL2_ARATH NITRILASE 2 >gi|322548|pir||S31969 nitrilase (EC 3.5.5.1) - Arabidopsis thaliana >gi|22656|emb|CAA48377| (X68305) nitrilase II [Arabidopsis thaliana] >gi|508733 (U09958) nitrilase [Arabidopsis thaliana] Length = 339
507	2028507	3' Pkc_Phospho_Site(41-43)
508	2028508	3' Pkc_Phospho_Site(11-13)
509	2028509	3' 1E-49 >gi|6166038|sp|P48421|CP83_ARATH CYTOCHROME P450 83A1 (CYPLXXXIII) >gi|2454176 (U69134) cytochrome P450 monooxygenase [Arabidopsis thaliana] >gi|3164128|dbj|BAA28532| (D78599) cytochrome P450 monooxygenase [Arabidopsis thaliana] >gi|4455306|emb|CAB36841.1| (AL035528) cytochrome P450 monooxygenase (CYP83A1) [Arabidopsis thaliana] Length = 502
510	2028510	3' Tyr_Phospho_Site(289-296)
511	2028511	3' Pkc_Phospho_Site(165-167)
512	2028512	5' Pkc_Phospho_Site(52-54)
513	2028513	5' 2E-28 >gi|5815233|gb|AAD52608.1|AF173378_1 (AF173378) 60S acidic ribosomal protein PO [Homo sapiens] Length = 239
514	2028514	5' Tyr_Phospho_Site(127-135)

515	2028515	9E-47 >emb|CAA05629.1| (AJ002597) membrane-associated salt-inducible protein like [Arabidopsis thaliana] Length = 428
516	2028516	Tyr_Phospho_Site(648-655)
517	2028517	2E-14 >gb|AAD17428| (AC006284) methyltransferase [Arabidopsis thaliana] Length = 619
518	2028518	6E-23 >dbj|BAA18924| (D61395) gamma-VPE [Arabidopsis thaliana] Length = 490
519	2028519	Pkc_Phospho_Site(79-81)
520	2028520	2E-28 >sp|P43601|YFJ1_YEAST HYPOTHETICAL 55.1 KD PROTEIN IN FAB1-PES4 INTERGENIC REGION >gi|1084743|pir||S56276 probable membrane protein YFR021w - yeast (Saccharomyces cerevisiae) >gi|836776|dbj|BAA09260.1| (D50617) YFR021W [Saccharomyces cerevisiae] Length = 500
521	2028521	3E-52 >sp|P46523|CLPA_BRANA ATP-DEPENDENT CLP PROTEASE ATP-BINDING SUBUNIT CLPA PRECURSOR >gi|480969|pir||S37557 clpA protein - rape (fragment) >gi|406311|emb|CAA53077| (X75328) clpA [Brassica napus] Length = 874
522	2028522	Tyr_Phospho_Site(1092-1098)
523	2028523	Tyr_Phospho_Site(727-735)
524	2028524	6E-64 >gb|AAD30599.1|AC007369_9 (AC007369) Similar to RNA helicases [Arabidopsis thaliana] Length = 1166
525	2028525	1E-106 >pir||S44943 sulfate adenylyltransferase (EC 2.7.7.4) - Arabidopsis thaliana >gi|2129743|pir||S68024 sulfate adenylyltransferase (EC 2.7.7.4) precursor (clone APS2) - Arabidopsis thaliana >gi|487404|emb|CAA55799| (X79210) sulfate adenylyltransferase [Arabidopsis thaliana] >gi|1228104 (U06276) ATP sulfurylase [Arabidopsis thaliana] >gi|1378028 (U40715) ATP sulfurylase precursor [Arabidopsis thaliana] >gi|1575324 (U59737) ATP sulfurylase [Arabidopsis thaliana] Length = 476
526	2028526	Tyr_Phospho_Site(1807-1814)
527	2028527	8E-59 >gi|3249077 (AC004473) Similar to prunasin hydrolase precursor gb|U50201 from Prunus serotina. ESTs gb|T21225 and gb|AA586305 come from this gene. [Arabidopsis thaliana] Length = 439
528	2028528	1E-69 >gb|AAD49995.1|AC007259_8 (AC007259) glucose transporter [Arabidopsis thaliana] Length = 522
529	2028529	4E-75 >gb|AAB63620.1| (AC002343) trehalase precusor isolog [Arabidopsis thaliana] Length = 557
530	2028530	2E-23 >gb|AAD21456.1| (AC007017) transcription factor E2F5 [Arabidopsis thaliana] Length = 532
531	2028531	Tyr_Phospho_Site(654-660)
532	2028532	4E-26 >gi|2494144 (AC002329) predicted leucine-rich protein [Arabidopsis thaliana] Length = 526
533	2028533	1E-13 >emb|CAA22523| (AL034563) transcription initiation factor iif, beta subunit [Schizosaccharomyces pombe] Length = 307
534	2028534	7E-12 >gb|AAD27870.1|AF134155_1 (AF134155) RING finger protein [Arabidopsis thaliana] Length = 170
535	2028535	Tyr_Phospho_Site(557-564)
536	2028536	3' Pkc_Phospho_Site(66-68)
537	2028537	3' 4E-17 >gi|2244792|emb|CAB10215.1| (Z97336) ankyrin like protein [Arabidopsis thaliana] Length = 936
538	2028538	3' Pkc_Phospho_Site(74-76)
539	2028539	3' Tyr_Phospho_Site(738-746)
540	2028540	3' Pkc_Phospho_Site(78-80)
541	2028541	3' Pkc_Phospho_Site(80-82)

542	2028542	5' 2E-66 >gi|2459443 (AC002332) NAD(P)-dependent cholesterol dehydrogenase [Arabidopsis thaliana] Length = 480
543	2028543	5' Tyr_Phospho_Site(543-551)
544	2028544	5' Tyr_Phospho_Site(245-252)
545	2028545	5' Pkc_Phospho_Site(1-3)
546	2028546	5' 6E-69 >gi|4538926|emb|CAB39662.1| (AL049483) phosphatidylserine decarboxylase [Arabidopsis thaliana] Length = 628
547	2028547	5' 3E-22 >gi|1931650 (U95973) disease resistance protein RPM1 isolog [Arabidopsis thaliana] Length = 821
548	2028548	1E-168 >emb|CAB52174.1| (AJ245407) syntaxin protein [Arabidopsis thaliana] Length = 341
549	2028549	Pkc_Phospho_Site(20-22)
550	2028550	3E-51 >gb|AAD50003.1|AC007259_16 (AC007259) Unknown protein [Arabidopsis thaliana] Length = 308
551	2028551	4E-53 >emb|CAB51834.1| (AJ243961) contains eukaryotic protein kinase domain PF|00069 [Oryza sativa] Length = 844
552	2028552	1E-62 >pir||S58494 IAA7 protein - Arabidopsis thaliana >gi|972917 (U18409) IAA7 [Arabidopsis thaliana] Length = 243
553	2028553	Pkc_Phospho_Site(14-16)
554	2028554	Tyr_Phospho_Site(164-171)
555	2028555	Pkc_Phospho_Site(31-33)
556	2028556	6E-33 >emb|CAB55281.1| (AL117212) WD domian, G-beta repeat protein [Schizosaccharomyces pombe] Length = 608
557	2028557	5E-98) >sp|O24496|GL2C_ARATH HYDROXYACYLGLUTATHIONE HYDROLASE CYTOPLASMIC (GLYOXALASE II) (GLX II) >gi|1924921|emb|CAA69644| (Y08357) hydroxyacylglutathione hydrolase [Arabidopsis thaliana] Length = 258
558	2028558	Pkc_Phospho_Site(64-66)
559	2028559	Tyr_Phospho_Site(250-256)
560	2028560	3' Tyr_Phospho_Site(168-174)
561	2028561	3' 2E-15 >gi|4539452|emb|CAB39932.1| (AL049500) phosphoribosylanthranilate transferase [Arabidopsis thaliana] Length = 857
562	2028562	3' 2E-19 >gi|2894378|emb|CAA74910.1| (Y14573) ribophorin I homologue [Hordeum vulgare] Length = 473
563	2028563	3' Pkc_Phospho_Site(39-41)
564	2028564	3' 4E-16 >gi|3913894|sp|O67825|IF2_AQUAE TRANSLATION INITIATION FACTOR IF-2 >gi|2984268 (AE000769) initiation factor IF-2 [Aquifex aeolicus] Length = 805
565	2028565	5' Pkc_Phospho_Site(231-233)
566	2028566	5' Pkc_Phospho_Site(4-6)
567	2028567	5' Tyr_Phospho_Site(17-24)
568	2028568	1E-60) >gi|2281109 (AC002333) endochitinase isolog [Arabidopsis thaliana] Length = 281
569	2028569	Pkc_Phospho_Site(61-63)
570	2028570	3E-19 >sp|P33174|KIF4_MOUSE KINESIN-LIKE PROTEIN KIF4 >gi|1083417|pir||A54803 microtubule-associated motor KIF4 - mouse >gi|563773|dbj|BAA02167| (D12646) KIF4 [Mus musculus] Length = 1231
571	2028571	6E-70 >gi|3367517 (AC004392) Similar to F4I1.26 beta-glucosidase gi|3128187 from A. thaliana BAC gb|AC004521. ESTs gb|N97083, gb|F19868 and gb|F15482 come from this gene. [Arabidopsis thaliana] Length = 527
572	2028572	Tyr_Phospho_Site(165-173)
573	2028573	Tyr_Phospho_Site(162-169)
574	2028574	4E-12 >emb|CAB38825.1| (AL035679) kinesin like protein [Arabidopsis

		thaliana] Length = 1121
575	2028575	9E-84 >gi|1931645 (U95973) Fe(II) transporter isolog [Arabidopsis thaliana] Length = 374
576	2028576	Tyr_Phospho_Site(299-305)
577	2028577	2E-66 >sp|O65355|GGH_ARATH GAMMA-GLUTAMYL HYDROLASE PRECURSOR (GAMMA-GLU-X CARBOXYPEPTIDASE) (CONJUGASE) (GH) >gi|3169656 (AF067141) gamma-glutamyl hydrolase [Arabidopsis thaliana] Length = 326
578	2028578	1E-39 >emb|CAB38294| (AL035605) formamidase-like protein [Arabidopsis thaliana] Length = 432
579	2028579	3' 3E-34 >gi|1707008 (U78721) 30S ribosomal protein S5 isolog [Arabidopsis thaliana] Length = 303
580	2028580	3' Rgd(732-734)
581	2028581	3' Pkc_Phospho_Site(28-30)
582	2028582	5' 9E-21 >gi|4263791|gb|AAD15451| (AC006068) receptor protein kinase [Arabidopsis thaliana] Length = 567
583	2028583	Tyr_Phospho_Site(710-718)
584	2028584	Tyr_Phospho_Site(632-638)
585	2028585	Pkc_Phospho_Site(77-79)
586	2028586	1E-63 >emb|CAA11285.1| (AJ223384) 26S proteasome regulatory ATPase subunit 10b (S10b) [Manduca sexta] Length = 396
587	2028587	Pkc_Phospho_Site(5-7)
588	2028588	Rgd(395-397)
589	2028589	2E-23 >gi|2149380 (U85036) syntaxin homolog [Arabidopsis thaliana] >gi|5281026|emb|CAB10553.2| (Z97344) syntaxin [Arabidopsis thaliana] Length = 255
590	2028590	Tyr_Phospho_Site(493-501)
591	2028591	Pkc_Phospho_Site(173-175)
592	2028592	8E-24 >emb|CAB07030| (Z92770) fadE2 [Mycobacterium tuberculosis] Length = 403
593	2028593	6E-25 >gi|3328893 (AE001320) Peptide Chain Release Factor 2 [Chlamydia trachomatis] Length = 369
594	2028594	Tyr_Phospho_Site(153-160)
595	2028595	Tyr_Phospho_Site(115-121)
596	2028596	Tyr_Phospho_Site(448-455)
597	2028597	Pkc_Phospho_Site(30-32)
598	2028598	Rgd(459-461)
599	2028599	3' 5E-21 >gi|1732517 (U62745) cytoskeletal protein [Arabidopsis thaliana] Length = 782
600	2028600	3' Pkc_Phospho_Site(330-332)
601	2028601	3' 8E-62 >gi|4097505 (U63020) D1 protein [Magnolia pyramidata] Length = 353
602	2028602	3' Tyr_Phospho_Site(118-126)
603	2028603	3' Pkc_Phospho_Site(36-38)
604	2028604	5' Tyr_Phospho_Site(720-727)
605	2028605	5' 2E-59 >gi|499301|emb|CAA54383| (X77116) ABI1 [Arabidopsis thaliana] >gi|549981 (U12856) abscisic acid insensitive protein [Arabidopsis thaliana] >gi|4538937|emb|CAB39673.1| (AL049483) protein phosphatase ABI1 [Arabidopsis thaliana] Length = 434
606	2028606	5' 6E-50 >gi|1709786|sp|P54904|PROC_ARATH PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE) >gi|541894|pir||JQ2334 pyrroline-5-carboxylate reductase (EC 1.5.1.2) - Arabidopsis thaliana >gi|166815 (M76538) pyrroline carboxylate reductase [Arabidopsis thaliana] >gi|1632776|emb|CAA70148|

607	2028607	5' Pkc_Phospho_Site(33-35)
608	2028608	1E-48 >gb|AAD10854.1| (U60135) serine/threonine protein phosphatase 2A-3 catalytic subunit [Arabidopsis thaliana] Length = 352
609	2028609	Pkc_Phospho_Site(56-58)
610	2028610	Tyr_Phospho_Site(62-68)
611	2028611	3E-17 >emb|CAB52561.1| (AL109819) stromal ascorbate peroxidase [Arabidopsis thaliana] Length = 372
612	2028612	8E-51) >gi|3421077 (AF043521) 20S proteasome subunit PAC1 [Arabidopsis thaliana] Length = 250
613	2028613	1E-82 >gi|3341695 (AC003672) thiamin pyrophosphokinase [Arabidopsis thaliana] Length = 263
614	2028614	Pkc_Phospho_Site(2-4)
615	2028615	1E-47 >emb|CAA18212.1| (AL022198) SERINE CARBOXYPEPTIDASE II-like protein [Arabidopsis thaliana] Length = 425
616	2028616	Pkc_Phospho_Site(55-57)
617	2028617	Pkc_Phospho_Site(15-17)
618	2028618	3E-27 >sp|P49691|RL4_ARATH 60S RIBOSOMAL PROTEIN L4 (L1) Length = 404
619	2028619	Pkc_Phospho_Site(42-44)
620	2028620	5E-27 >gi|3252815 (AC004705) vacuolar sorting receptor-like protein [Arabidopsis thaliana] >gi|3810588 (AC005398) vacuolar sorting receptor-like protein [Arabidopsis thaliana] Length = 628
621	2028621	2E-43 >emb|CAA23023.1| (AL035394) phosphatase like protein [Arabidopsis thaliana] Length = 350
622	2028622	3' Pkc_Phospho_Site(55-57)
623	2028623	5' Pkc_Phospho_Site(4-6)
624	2028624	5' Pkc_Phospho_Site(9-11)
625	2028625	5' Tyr_Phospho_Site(35-41)
626	2028626	4E-34 >gi|3859659|emb|CAA20566.1| (AL031394) potassium transporter AtKT5p (AtKT5) [Arabidopsis thaliana] Length = 846
627	2028627	5' 3E-74 >gi|585421|sp|P38418|LOXC_ARATH LIPOXYGENASE, CHLOROPLAST PRECURSOR >gi|541879|pir||JQ2391 lipoxygenase (EC 1.13.11.12) AtLox2 - Arabidopsis thaliana >gi|431258 (L23968) lipoxygenase [Arabidopsis thaliana] Length = 896
628	2028628	Tyr_Phospho_Site(35-41)
629	2028629	2E-29 >gi|2621798 (AE000850) transcriptional regulator [Methanobacterium thermoautotrophicum] Length = 151
630	2028630	2E-53 >gi|1181531 (L41244) thionin [Arabidopsis thaliana] >gi|1586833|prf||2204399A thionin [Arabidopsis thaliana] Length = 134
631	2028631	2E-34 >gb|AAC69619.1| (AF072736) beta-glucosidase [Pinus contorta] Length = 513
632	2028632	7E-32 >gi|3599491 (AF085149) aminotransferase [Capsicum chinense] Length = 459
633	2028633	Pkc_Phospho_Site(39-41)
634	2028634	Pkc_Phospho_Site(23-25)
635	2028635	Tyr_Phospho_Site(92-99)
636	2028636	1E-82 >emb|CAA11525.1| (AJ223635) transcription factor IIA large subunit [Arabidopsis thaliana] Length = 375
637	2028637	7E-27 >pir||S30578 proteinase inhibitor II - Arabidopsis thaliana >gi|16427|emb|CAA48892| (X69139) protease inhibitor II [Arabidopsis thaliana] >gi|4038041 (AC005936) proteinase inhibitor II [Arabidopsis thaliana] Length = 77
638	2028638	2E-68 >dbj|BAA19751| (D85339) hydroxypyruvate reductase [Arabidopsis thaliana] Length = 386

639	2028639	7E-12 >sp|O07051|LTAA_AERJA L-ALLO-THREONINE ALDOLASE (L-ALLO-TA) (L-ALLO-THREONINE ACETALDEHYDE-LYASE) >gi|2190272|dbj|BAA20404| (D87890) L-allo-threonine aldolase [Aeromonas jandaei] Length = 338
640	2028640	1E-12 >gi|3193298 (AF069298) T14P8.17 gene product [Arabidopsis thaliana] Length = 154
641	2028641	Tyr_Phospho_Site(213-220)
642	2028642	4E-25 >sp|O49972|DCA2_BRAJU S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME 2 (ADOMETDC 2) (SAMDC 2) >gi|2662406 (U80916) S-adenosyl-L-methionine decarboxylase [Brassica juncea] Length = 369
643	2028643	3' 2E-13 >gi|2641638 (AF032883) AtJ3 [Arabidopsis thaliana] Length = 420
644	2028644	3' Tyr_Phospho_Site(296-303)
645	2028645	5' Tyr_Phospho_Site(29-36)
646	2028646	5' 1E-73 >gi|5902365|gb|AAD55467.1|AC009322_7 (AC009322) splicing factor Prp8 [Arabidopsis thaliana] Length = 2359
647	2028647	5' 6E-37 >gi|1542941|emb|CAA55006| (X78116) Acetoacetyl-coenzyme A thiolase [Raphanus sativus] Length = 406
648	2028648	Rgd(383-385)
649	2028649	4E-61 >gb|AAD45605.1|AF160729_1 (AF160729) isovaleryl-CoA-dehydrogenase precursor [Arabidopsis thaliana] Length = 409
650	2028650	Tyr_Phospho_Site(1078-1085)
651	2028651	1E-105 >emb|CAA16684| (AL021684) oxoglutarate dehydrogenase - like protein [Arabidopsis thaliana] Length = 973
652	2028652	2E-52 >sp|Q45223|HBD_BRAJA 3-HYDROXYBUTYRYL-COA DEHYDROGENASE (BETA-HYDROXYBUTYRYL-COA DEHYDROGENASE) (BHBD) >gi|1209052 (U32229) HbdA [Bradyrhizobium japonicum] Length = 293
653	2028653	Tyr_Phospho_Site(711-719)
654	2028654	9E-24 >gi|3738320 (AC005170) cinnamoyl CoA reductase [Arabidopsis thaliana] Length = 303
655	2028655	3E-67) >gi|2952433 (AF051135) ubiquitin activating enzyme E1 [Arabidopsis thaliana] Length = 454
656	2028656	Pkc_Phospho_Site(31-33)
657	2028657	Tonb_Dependent_Rec_1(1-75)
658	2028658	8E-87 >sp|P49661|COPD_ORYSA COATOMER DELTA SUBUNIT (DELTA-COAT PROTEIN) (DELTA-COP) (ARCHAIN) >gi|1314049|emb|CAA91901| (Z67962) archain/delta-COP [Oryza sativa] Length = 518
659	2028659	9E-40 >dbj|BAA84386.1| (AP000423) ycf3 [Arabidopsis thaliana] Length = 126
660	2028660	3' Pkc_Phospho_Site(8-10)
661	2028661	3' Pkc_Phospho_Site(123-125)
662	2028662	5' Zinc_Finger_C2h2(63-85)
663	2028663	5' Tyr_Phospho_Site(544-551)
664	2028664	5' 2E-54 >gi|2129727|pir||S71229 RNA-binding protein 37 - Arabidopsis thaliana >gi|1174153 (U44134) RNA-binding protein [Arabidopsis thaliana] Length = 336
665	2028665	Tyr_Phospho_Site(383-390)
666	2028666	1E-51 >emb|CAA17552| (AL021961) Phosphoglycerate dehydrogenase - like protein [Arabidopsis thaliana] Length = 603
667	2028667	Tyr_Phospho_Site(371-378)
668	2028668	Pkc_Phospho_Site(25-27)

669	2028669	Pkc_Phospho_Site(14-16)
670	2028670	3E-29 >gi|4155557 (AE001526) CYCLOPOCYCLOPROPANE FATTY ACID SYNTHASE [Helicobacter pylori J99] Length = 389
671	2028671	2E-79 >emb|CAA09208| (AJ010469) RNA helicase [Arabidopsis thaliana] Length = 360
672	2028672	Tyr_Phospho_Site(319-325)
673	2028673	1E-113 >gb|AAD55787.1|AF181966_1 (AF181966) methylenetetrahydrofolate reductase MTHFR1 [Arabidopsis thaliana] Length = 592
674	2028674	Tyr_Phospho_Site(1157-1163)
675	2028675	3E-69) >gi|3421090 (AF043525) 20S proteasome subunit PAE2 [Arabidopsis thaliana] Length = 237
676	2028676	1E-56 >gi|4063738 (AC005851) zinc finger protein [Arabidopsis thaliana] >gi|4803961|gb|AAD29833.1|AC006202_11 (AC006202) unknown protein [Arabidopsis thaliana] Length = 284
677	2028677	Pkc_Phospho_Site(22-24)
678	2028678	Tyr_Phospho_Site(174-180)
679	2028679	4E-43 >emb|CAA47807| (X67421) extA [Arabidopsis thaliana] Length = 127
680	2028680	3' Tyr_Phospho_Site(195-202)
681	2028681	3' 4E-14 >gi|120532|sp|P19976|FRI_SOYBN FERRITIN PRECURSOR (SOF-35) >gi|81773|pir||A40992 ferritin precursor - soybean >gi|169953 (M64337) ferritin light chain [Glycine max] Length = 250
682	2028682	3' Rgd(36-38)
683	2028683	3' 4E-35 >gi|3047064 (AF058825) contains similarity to peptidyl-prolyl cis-trans isomerase (Pfam: pro_isomerase.hmm, score: 23.86 and 28.41 [Arabidopsis thaliana] Length = 281
684	2028684	3' Pkc_Phospho_Site(11-13)
685	2028685	5' Pkc_Phospho_Site(47-49)
686	2028686	5' 2E-19 >gi|6322411|ref|NP_012485.1|MTR4| RNA helicase; Mtr4p >gi|1352980|sp|P47047|MTR4_YEAST ATP-DEPENDENT RNA HELICASE DOB1 (MRNA TRANSPORT REGULATOR MTR4) >gi|1078374|pir||S56822 SKI2 protein homolog YJL050w - yeast (Saccharomyces cerevisiae) >gi|1008185|emb|CAA89341| (Z49325) ORF YJL050w
687	2028687	5' Tyr_Phospho_Site(622-629)
688	2028688	5' Rgd(156-158)
689	2028689	Pkc_Phospho_Site(29-31)
690	2028690	Tyr_Phospho_Site(350-356)
691	2028691	1E-14 >gi|3834312 (AC005679) Strong similarity to glycoprotein EP1 gb|L16983 Daucus carota and a member of S locus glycoprotein family PF|00954. ESTs gb|AA067487, gb|Z35737, gb|Z30815, gb|Z35350, gb|AA713171, gb|AI100553, gb|Z34248, gb|AA728536, gb|Z30816 an... Length
692	2028692	Pkc_Phospho_Site(2-4)
693	2028693	6E-28 >gi|4102703 (AF015274) ribulose-5-phosphate-3-epimerase [Arabidopsis thaliana] Length = 281
694	2028694	Tyr_Phospho_Site(295-303)
695	2028695	Tyr_Phospho_Site(790-796)
696	2028696	Tyr_Phospho_Site(151-158)
697	2028697	Pkc_Phospho_Site(26-28)
698	2028698	1E-59 >emb|CAA74372| (Y14044) geranylgeranyl reductase [Arabidopsis thaliana] Length = 472
699	2028699	Tyr_Phospho_Site(823-830)
700	2028700	Tyr_Phospho_Site(159-166)
701	2028701	2E-13 >gi|4249409 (AC006072) sugar transporter [Arabidopsis

		thaliana] Length = 348
702	2028702	8E-76 >emb|CAB38611.1| (AL035656) extensin-like protein [Arabidopsis thaliana] Length = 448
703	2028703	6E-83) >sp|P29513|TBB5_ARATH TUBULIN BETA-5 CHAIN >gi|320186|pir||JQ1589 tubulin beta-5 chain - Arabidopsis thaliana >gi|166902 (M84702) beta-5 tubulin [Arabidopsis thaliana] Length = 449
704	2028704	3' Tyr_Phospho_Site(418-424)
705	2028705	3' 4E-39 >gi|4103987 (AF030516) 5,10-methylenetetrahydrofolate dehydrogenase-5,10-methenyltetrahydrofolate cyclohydrolase [Pisum sativum] >gi|6002383|emb|CAB56756.1| (AJ011589) 5,10-methylenetetrahydrofolate dehydrogenase: 5,10-methenyltetrahydrofolate cyclohydrolase [Pisum sativum] Length = 294
706	2028706	3' Tyr_Phospho_Site(470-478)
707	2028707	5' Pkc_Phospho_Site(35-37)
708	2028708	5' Pkc_Phospho_Site(18-20)
709	2028709	5' Pkc_Phospho_Site(236-238)
710	2028710	5' Pkc_Phospho_Site(7-9)
711	2028711	5' 6E-43 >gi|6006879|gb|AAF00654.1|AC008153_6 (AC008153) eukaryotic translation initiation factor 3 subunit [Arabidopsis thaliana] Length = 294
712	2028712	5' 2E-61 >gi|1750376 (U80808) ubiquitin activating enzyme [Arabidopsis thaliana] >gi|3150409 (AC004165) ubiquitin activating enzyme (UBA1) [Arabidopsis thaliana] Length = 1080
713	2028713	5' Tyr_Phospho_Site(141-148)
714	2028714	5' Pkc_Phospho_Site(186-188)
715	2028715	5' 3E-29 >gi|3914191|sp|P56558|OGT1_RAT UDP-N-ACETYLGLUCOSAMINE—PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 110 KD SUBUNIT (O-GLCNAC TRANSFERASE P110 SUBUNIT) >gi|1931579 (U76557) O-GlcNAc transferase, p110 subunit [Rattus norvegicus] Length = 1036
716	2028716	5' 3E-71 >gi|5931694|emb|CAB56597.1| (Y18470) Exportin1 (XPO1) protein [Arabidopsis thaliana] Length = 1075
717	2028717	Tyr_Phospho_Site(450-458)
718	2028718	5E-43 >pir||S58118 thioredoxin - Arabidopsis thaliana >gi|992962|emb|CAA84611| (Z35474) thioredoxin [Arabidopsis thaliana] >gi|1388076 (U35640) thioredoxin h [Arabidopsis thaliana] Length = 118
719	2028719	9E-45 >gi|3287677 (AC003979) Contains similarity to transcription factor (TINY) isolog T02O04.22 gb|2062174 from A. thaliana BAC gb|AC001645. [Arabidopsis thaliana] Length = 144
720	2028720	2E-11 >emb|CAB45279.1| (AL079313) hypothetical protein, similar to (M97204) goliath protein [Drosophila melanogaster] [Homo sapiens] Length = 104
721	2028721	1E-94 >gb|AAD20931| (AC006234) diacylglycerol kinase [Arabidopsis thaliana] Length = 493
722	2028722	Tyr_Phospho_Site(688-695)
723	2028723	Pkc_Phospho_Site(45-47)
724	2028724	Tyr_Phospho_Site(303-311)
725	2028725	1E-178 >gi|4220485 (AC006069) beta-1,3-glucanase [Arabidopsis thaliana] Length = 439
726	2028726	2E-32 >sp|P34124|PRS8_DICDI 26S PROTEASE REGULATORY SUBUNIT 8 (TAT-BINDING PROTEIN HOMOLOG 10) >gi|422297|pir||JN0610 probable transcription factor DdTBP10 - slime mold (Dictyostelium discoideum) (fragment) >gi|290057 (L16579) HIV1 TAT-binding protein [Dictyostelium discoideum] Length = 389
727	2028727	7E-86 >gb|AAD25787.1|AC006577_23 (AC006577) Similar to gi|1653162 (p)ppGpp 3-pyrophosphohydrolase from Synechocystis sp genome gb|D90911.

		EST gb|W43807 comes from this gene. [Arabidopsis thaliana] Length = 715
728	2028728	3E-13 >gi|3420745 (AF079445) TipC [Dictyostelium discoideum] Length = 3848
729	2028729	3' 2E-16 >gi|4538906|emb|CAB39643.1| (AL049482) choline kinase GmCK2p-like protein [Arabidopsis thaliana] Length = 346
730	2028730	3' Pkc_Phospho_Site(64-66)
731	2028731	3' Pkc_Phospho_Site(114-116)
732	2028732	3' Tyr_Phospho_Site(227-234)
733	2028733	3' Rgd(568-570)
734	2028734	3' Tyr_Phospho_Site(13-20)
735	2028735	3' Tyr_Phospho_Site(172-180)
736	2028736	5' 4E-64 >gi|2129613|pir||A57632 homeotic protein BEL1 - Arabidopsis thaliana >gi|1122533 (U39944) BELL1 [Arabidopsis thaliana] Length = 610
737	2028737	5' 2E-21 >gi|3912917|gb|AAC78693.1| (AF001308) NAK-like ser/thr protein kinase [Arabidopsis thaliana] Length = 707
738	2028738	5' Pkc_Phospho_Site(3-5)
739	2028739	5' Tyr_Phospho_Site(301-309)
740	2028740	Pkc_Phospho_Site(69-71)
741	2028741	Pkc_Phospho_Site(38-40)
742	2028742	Tyr_Phospho_Site(478-485)
743	2028743	Pkc_Phospho_Site(2-4)
744	2028744	1E-31 >emb|CAA16524.1| (AL021633) DNA topoisomerase like-protein [Arabidopsis thaliana] Length = 1179
745	2028745	1E-71) >gi|2347191 (AC002338) DNA binding protein isolog [Arabidopsis thaliana] >gi|3150397 (AC004165) DNA-binding protein [Arabidopsis thaliana] Length = 393
746	2028746	2E-80 >gi|3377808 (AF075597) contains similarity to Nicotiana alata pistil extensin-like protein (GB:U45958) [Arabidopsis thaliana] Length = 165
747	2028747	1E-33 >sp|P54888|P5C2_ARATH DELTA 1-PYRROLINE-5-CARBOXYLATE SYNTHETASE B (P5CS B) [INCLUDES: GLUTAMATE 5-KINASE (GAMMA-GLUTAMYL KINASE) (GK); GAMMA-GLUTAMYL PHOSPHATE REDUCTASE (GPR) (GLUTAMATE-5-SEMIALDEHYDE DEHYDROGENASE) (GLUTAMYL-GAMMA-SEMIALDE... >gi|887388|emb|CAA60447| (X86778) pyrroline-5-carboxylate synthetase B [Arabidopsis thaliana] >gi|1669658|emb|CAA70527| (Y09355) pyrroline-5-carboxlyate synthetase [Arabidopsis thaliana] Length = 726
748	2028748	2E-54 >emb|CAB45881.1| (AL080282) berberine bridge enzyme-like protein [Arabidopsis thaliana] Length = 530
749	2028749	5E-47 >gb|AAD39281.1|AC007576_4 (AC007576) initiation factor 5A-4 [Arabidopsis thaliana] Length = 158
750	2028750	3E-43 >gi|3941522 (AF062915) transcription factor [Arabidopsis thaliana] Length = 249
751	2028751	8E-19 >emb|CAB10269.1| (Z97337) hydroxyproline-rich glycoprotein homolog [Arabidopsis thaliana] Length = 507
752	2028752	Tyr_Phospho_Site(757-764)
753	2028753	Tyr_Phospho_Site(316-322)
754	2028754	3' Tyr_Phospho_Site(427-434)
755	2028755	3' Tyr_Phospho_Site(730-738)
756	2028756	3' 8E-32 >gi|1076534|pir||A55333 monodehydroascorbate reductase (NADH) (EC 1.6.5.4) - garden pea >gi|497120 (U06461) monodehydroascorbate reductase [Pisum sativum] Length = 433
757	2028757	5' 1E-16 >gi|3337095|dbj|BAA31843| (AB016206) polygalacturonase inhibitor (PGIP) [Citrus iyo] Length = 327
758	2028758	5' 4E-42 >gi|4586249|emb|CAB40990.1| (AL049640) pollen surface protein [Arabidopsis thaliana] Length = 403

759	2028759	5' Pkc_Phospho_Site(5-7)
760	2028760	5' Tyr_Phospho_Site(560-566)
761	2028761	2E-40 >emb|CAA76178.1| (Y16327) cyclic nucleotide-regulated ion channel [Arabidopsis thaliana] Length = 716
762	2028762	Tyr_Phospho_Site(178-185)
763	2028763	7E-12 >dbj|BAA13831| (D89169) similar to Saccharomyces cerevisiae SCD6 protein, SWISS-PROT Accession Number P45978 [Schizosaccharomyces pombe] Length = 370
764	2028764	Rgd(288-290)
765	2028765	Tyr_Phospho_Site(21-27)
766	2028766	Tyr_Phospho_Site(722-729)
767	2028767	Tyr_Phospho_Site(1033-1039)
768	2028768	Pkc_Phospho_Site(45-47)
769	2028769	4E-95 >gi|4090884 (AF025333) vesicle-associated membrane protein 7B; synaptobrevin 7B [Arabidopsis thaliana] Length = 219
770	2028770	3E-82 >emb|CAA10320| (AJ131205) mitochondrial NAD-dependent malate dehydrogenase [Arabidopsis thaliana] Length = 341
771	2028771	Pkc_Phospho_Site(277-279)
772	2028772	Pkc_Phospho_Site(13-15)
773	2028773	Pkc_Phospho_Site(168-170)
774	2028774	6E-39 >emb|CAB55622.1| (AJ011044) cysteine synthase [Arabidopsis thaliana] Length = 176
775	2028775	1E-27 >pir||S65071 cystatin - field mustard >gi|762785 (L41355) cysteine proteinase inhibitor [Brassica campestris] Length = 199
776	2028776	6E-62 >gi|3201633 (AC004669) cell division protein [Arabidopsis thaliana] Length = 695
777	2028777	5E-81) >sp|P25069|CAL2_ARATH CALMODULIN-2/3/5 >gi|99671|pir||S22503 calmodulin - Arabidopsis thaliana >gi|1076437|pir||S53006 calmodulin - leaf mustard >gi|2146726|pir||S71513 calmodulin - Arabidopsis thaliana >gi|166651 (M38380) calmodulin-2 [Arabidopsis thaliana] >gi|166653 (M73711) calmodulin-3 [Arabidopsis thaliana] >gi|474183|emb|CAA47690| (X67273) calmodulin [Arabidopsis thaliana] >gi|497992 (U10150) calmodulin [Brassica napus] >gi|899058 (M88307) calmodulin [Brassica juncea] >gi|1183005|dbj|BAA08283| (D45848) calmodulin [Arabidopsis thaliana] >gi|3402706 (AC004261) unknown protein [Arabidopsis thaliana] >gi|3885333 (AC005623) calmodulin [Arabidopsis thaliana] >gi|228407|prf||1803520A calmodulin 2 [Arabidopsis thaliana] Length = 149
778	2028778	1E-13 >emb|CAA18500| (AL022373) Myc-type transcription factor [Arabidopsis thaliana] Length = 272
779	2028779	3' Pkc_Phospho_Site(37-39)
780	2028780	5' Pkc_Phospho_Site(59-61)
781	2028781	Tyr_Phospho_Site(305-312)
782	2028782	Tyr_Phospho_Site(2-9)
783	2028783	Pkc_Phospho_Site(63-65)
784	2028784	Pkc_Phospho_Site(87-89)
785	2028785	Tyr_Phospho_Site(412-419)
786	2028786	4E-39 >gb|AAD46410.1|AF096260_1 (AF096260) ER66 protein [Lycopersicon esculentum] Length = 558
787	2028787	Pkc_Phospho_Site(21-23)
788	2028788	Pkc_Phospho_Site(24-26)
789	2028789	Tyr_Phospho_Site(68-75)
790	2028790	3' 4E-27 >gi|4678261|emb|CAB41122.1| (AL049657) proteasome regulatory subunit [Arabidopsis thaliana] Length = 406
791	2028791	3' Pkc_Phospho_Site(129-131)

792	2028792	5' Tyr_Phospho_Site(6-12)
793	2028793	5' 9E-27 >gi|4914387|gb|AAD32922.1|AC007167_4 (AC007167) heat-shock protein [Arabidopsis thaliana] Length = 780
794	2028794	Serpin(210-220)
795	2028795	Tyr_Phospho_Site(327-334)
796	2028796	Pkc_Phospho_Site(35-37)
797	2028797	1E-45 >gi|4093155 (AF088281) phytochrome-associated protein 1 [Arabidopsis thaliana] Length = 267
798	2028798	3E-51 >gb|AAD25794.1|AC006550_2 (AC006550) Similar to gb|U51990 pre-mRNA-splicing factor hPrp18 from Homo sapiens. ESTs gb|T46391 and gb|AA721815 come from this gene. [Arabidopsis thaliana] Length = 420
799	2028799	Pkc_Phospho_Site(11-13)
800	2028800	Tyr_Phospho_Site(202-209)
801	2028801	3E-48 >emb|CAB39679.1| (AL049483) beta-galactosidase [Arabidopsis thaliana] Length = 729
802	2028802	4E-47 >emb|CAA18465.1| (AL022347) serine/threonine kinase-like protein [Arabidopsis thaliana] Length = 633
803	2028803	1E-74 >gi|3044218 (AF057144) signal peptidase [Arabidopsis thaliana] Length = 167
804	2028804	Tyr_Phospho_Site(707-715)
805	2028805	Pkc_Phospho_Site(22-24)
806	2028806	Tyr_Phospho_Site(325-332)
807	2028807	5E-65 >emb|CAB16773.1| (Z99707) Cu2+-transporting ATPase-like protein [Arabidopsis thaliana] Length = 819
808	2028808	8E-63 >gb|AAD17333| (AF125574) lysyl-tRNA synthetase; LysRS [Arabidopsis thaliana] >gi|6041823|gb|AAF02138.1|AC009918_10 (AC009918) lysyl-tRNA synthetase [Arabidopsis thaliana] Length = 626
809	2028809	2E-56 >gi|2909781 (AF020288) MgATP-energized glutathione S-conjugate pump [Arabidopsis thaliana] Length = 1623
810	2028810	3' Tyr_Phospho_Site(749-756)
811	2028811	3' 2E-20 >gi|1655424|dbj|BAA11944| (D83531) GDP dissociation inhibitor [Arabidopsis thaliana] >gi|3212878 (AC004005) GDP dissociation inhibitor [Arabidopsis thaliana] Length = 445
812	2028812	3' Tyr_Phospho_Site(244-251)
813	2028813	3' 6E-15 >gi|4325346|gb|AAD17345.1| (AF128393) similar to N-ethylmaleimide sensitive fusion proteins; contains similarity to ATPases (Pfam: PF00004, Score=307.7, E=1.4e-88n N=1) [Arabidopsis thaliana] Length = 772
814	2028814	3' Rgd(690-692)
815	2028815	5' 1E-33 >gi|2905657 (AF047469) arsenite translocating ATPase [Homo sapiens] Length = 348
816	2028816	5' Tyr_Phospho_Site(417-424)
817	2028817	5' 5E-44 >gi|5929906|gb|AAD56636.1|AF162150_1 (AF162150) COP1-interacting protein CIP8 [Arabidopsis thaliana] Length = 334
818	2028818	Tyr_Phospho_Site(654-662)
819	2028819	Tyr_Phospho_Site(556-564)
820	2028820	Pkc_Phospho_Site(13-15)
821	2028821	4E-35 >sp|P49688|RS2_ARATH 40S RIBOSOMAL PROTEIN S2 >gi|2335095 (AC002339) 40S ribosomal protein S2 [Arabidopsis thaliana] Length = 285
822	2028822	6E-23 >ref|NP_004862.1|PGOSR1| golgi SNAP receptor complex member 1 >gi|4234774 (AF073926) cis-Golgi SNARE p28 [Homo sapiens] Length = 250
823	2028823	Tyr_Phospho_Site(409-416)

824	2028824	Pkc_Phospho_Site(2-4)
825	2028825	5E-66) >emb|CAB16796.1| (Z99707) MAP3K-like protein kinase [Arabidopsis thaliana] Length = 799
826	2028826	7E-71 >emb|CAB10557.1| (Z97344) trehalose-6-phosphate synthase like protein [Arabidopsis thaliana] Length = 865
827	2028827	1E-139 >gi|2262167 (AC002329) cytosolic ribosomal protein S4 [Arabidopsis thaliana] Length = 261
828	2028828	4E-12 >gi|3327957 (AF060490) TLS-associated protein TASR-2 [Mus musculus] >gi|3327976 (AF067730) TLS-associated protein TASR-2 [Homo sapiens] Length = 262
829	2028829	2E-30 >pir||S59544 stress-induced protein OZI1 precursor - Arabidopsis thaliana >gi|790583 (U20347) mRNA corresponding to this gene accumulates in response to ozone stress and pathogen (bacterial) infection; pathogenesis-related protein [Arabidopsis thaliana] >gi|2252869 (AF013294) No definition line found [Arabidopsis thaliana] Length = 80
830	2028830	5E-47 >dbj|BAA24694| (D88206) protein kinase [Arabidopsis thaliana] Length = 426
831	2028831	Pkc_Phospho_Site(8-10)
832	2028832	Tyr_Phospho_Site(58-64)
833	2028833	3' 1E-14 >gi|4027895 (AF049352) alpha-expansin precursor [Nicotiana tabacum] Length = 257
834	2028834	5' Tyr_Phospho_Site(166-173)
835	2028835	5' 8E-44 >gi|484656|pir||JU0182 monodehydroascorbate reductase (NADH) (EC 1.6.5.4) - cucumber >gi|452165|dbj|BAA05408| (D26392) monodehydroascorbate reductase [Cucumis sativus] Length = 434
836	2028836	Tyr_Phospho_Site(419-426)
837	2028837	Tyr_Phospho_Site(579-585)
838	2028838	3E-32 >sp|Q45223|HBD_BRAJA 3-HYDROXYBUTYRYL-COA DEHYDROGENASE (BETA-HYDROXYBUTYRYL-COA DEHYDROGENASE) (BHBD) >gi|1209052 (U32229) HbdA [Bradyrhizobium japonicum] Length = 293
839	2028839	1E-14 >gi|3461840 (AC005315) reverse transcriptase [Arabidopsis thaliana] Length = 1529
840	2028840	1E-16 >dbj|BAA75684.1| (AB017693) transcription factor [Nicotiana tabacum] Length = 291
841	2028841	8E-66 >gi|2160694 (U73528) B' regulatory subunit of PP2A [Arabidopsis thaliana] Length = 522
842	2028842	Tyr_Phospho_Site(194-200)
843	2028843	Pkc_Phospho_Site(28-30)
844	2028844	3' 2E-23 >gi|2129770|pir||S71224 xyloglucan endotransglycosylase-related protein XTR-2 - Arabidopsis thaliana >gi|1244756 (U43487) xyloglucan endotransglycosylase-related protein [Arabidopsis thaliana] >gi|2154611|dbj|BAA20290| (D63510) endoxyloglucan transferase related protein [Arabidopsis thaliana] >gi|5533311|gb|AAD45124.1|AF163820_1 (AF163820) endoxyloglucan transferase [Arabidopsis thaliana] Length = 332
845	2028845	3' Pkc_Phospho_Site(42-44)
846	2028846	3' Tyr_Phospho_Site(152-160)
847	2028847	3' 8E-13 >gi|1076421|pir||S46523 transcription factor TGA3 - Arabidopsis thaliana >gi|304113 (L10209) transcription factor [Arabidopsis thaliana] Length = 384
848	2028848	3' Pkc_Phospho_Site(2-4)
849	2028849	5' Tyr_Phospho_Site(764-771)
850	2028850	2E-65 >emb|CAA67338| (X98806) peroxidase ATP20a [Arabidopsis thaliana] Length = 330
851	2028851	3E-99 >emb|CAB45075.1| (AL078637) serine/threonine kinase-like protein [Arabidopsis thaliana] Length = 445

852	2028852	4E-71 >emb|CAB10698| (Z97558) argininosuccinate lyase [Arabidopsis thaliana] Length = 517
853	2028853	4E-72 >sp|P46086|KIME_ARATH MEVALONATE KINASE (MK) >gi|541880|pir||S42088 mevalonate kinase (EC 2.7.1.36) - Arabidopsis thaliana >gi|456614|emb|CAA54820| (X77793) mevalonate kinase [Arabidopsis thaliana] >gi|4883990|gb|AAD31719.1|AF141853_1 (AF141853) mevalonate kinase [Arabidopsis thaliana] Length = 378
854	2028854	Tyr_Phospho_Site(53-60)
855	2028855	Pkc_Phospho_Site(62-64)
856	2028856	2E-17 >gi|1899188 (U90212) DNA binding protein ACBF [Nicotiana tabacum] Length = 428
857	2028857	Tyr_Phospho_Site(364-371)
858	2028858	2E-65 >sp|P25284|NUEM_NEUCR NADH-UBIQUINONE OXIDOREDUCTASE 40 KD SUBUNIT PRECURSOR (COMPLEX I-40KD) (CI-40KD) >gi|101865|pir||S13025 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) 40K chain - Neurospora crassa >gi|3046|emb|CAA39
859	2028859	7E-25 >gi|2191150 (AF007269) similar to mitochondrial carrier family [Arabidopsis thaliana] Length = 352
860	2028860	3' 8E-21 >gi|4218120|emb|CAA22974.1| (AL035353) Proline-rich APG-like protein [Arabidopsis thaliana] Length = 367
861	2028861	3' Tyr_Phospho_Site(684-690)
862	2028862	3' Pkc_Phospho_Site(49-51)
863	2028863	3' Tyr_Phospho_Site(485-493)
864	2028864	3' Wd_Repeats(436-450)
865	2028865	3' Pkc_Phospho_Site(50-52)
866	2028866	3' Pkc_Phospho_Site(23-25)
867	2028867	3' Pkc_Phospho_Site(2-4)
868	2028868	3' Pkc_Phospho_Site(5-7)
869	2028869	5' 3E-65 >gi|2827708|emb|CAA16681| (AL021684) myb - related protein [Arabidopsis thaliana] Length = 374
870	2028870	5' Pkc_Phospho_Site(101-103)
871	2028871	5' 7E-22 >gi|322752|pir||A44226 auxin-independent growth promoter - Nicotiana tabacum >gi|559921|emb|CAA56570| (X80301) axi 1 [Nicotiana tabacum] Length = 569
872	2028872	Pkc_Phospho_Site(26-28)
873	2028873	4E-40 >gi|2435517 (AF024504) contains similarity to peptidase family A1 [Arabidopsis thaliana] Length = 472
874	2028874	Pkc_Phospho_Site(47-49)
875	2028875	6E-43 >emb|CAB43855.1| (AL078465) isp4 like protein [Arabidopsis thaliana] Length = 753
876	2028876	2E-53 >sp|P54641|VATX_DICDI VACUOLAR ATP SYNTHASE SUBUNIT AC39 (V-ATPASE AC39 SUBUNIT) (41 KD ACCESSORY PROTEIN) (DVA41) >gi|626048|pir||A55016 lysosomal membrane protein DVA41 - slime mold (Dictyostelium discoideum) >gi|532733 (U13150) vacuolar ATPase subunit DVA41 [Dictyosteli
877	2028877	1E-36 >emb|CAA18734.1| (AL022604) cysteine proteinase-like protein [Arabidopsis thaliana] Length = 355
878	2028878	8E-12 >pir||S71365 AP2 domain-containing protein - Arabidopsis thaliana >gi|1209099 (U40256) AINTEGUMENTA [Arabidopsis thaliana] >gi|1244708 (U41339) ANT [Arabidopsis thaliana] >gi|4490720|emb|CAB38923.1| (AL035709) ovule development protein aintegumenta (ANT) [Arabidopsis thaliana] Length = 555
879	2028879	2E-60 >gi|3738302 (AC005309) tubby-like protein [Arabidopsis thaliana] >gi|4249398 (AC006072) tubby protein [Arabidopsis thaliana] Length = 407

880	2028880	Pkc_Phospho_Site(29-31)
881	2028881	1E-67) >emb|CAA16700.1| (AL021687) kinase-like protein [Arabidopsis thaliana] Length = 290
882	2028882	Pkc_Phospho_Site(2-4)
883	2028883	1E-23 >emb|CAB40952.1| (AL049638) C-4 sterol methyl oxidase [Arabidopsis thaliana] Length = 303
884	2028884	3' Pkc_Phospho_Site(23-25)
885	2028885	3' Pkc_Phospho_Site(11-13)
886	2028886	3' 7E-12 >gi|5103828|gb|AAD39658.1|AC007591_23 (AC007591) Similar to gi|22113 Ac transposase (ORFa) from Zea mays transcript gb|X05424. [Arabidopsis thaliana] Length = 799
887	2028887	3' Pkc_Phospho_Site(116-118)
888	2028888	3' Tyr_Phospho_Site(532-539)
889	2028889	5' Pkc_Phospho_Site(61-63)
890	2028890	5' Pkc_Phospho_Site(137-139)
891	2028891	5' Pkc_Phospho_Site(26-28)
892	2028892	5' Pkc_Phospho_Site(74-76)
893	2028893	5' Tyr_Phospho_Site(604-610)
894	2028894	5' Tyr_Phospho_Site(666-674)
895	2028895	8E-51 >gi|1336084 (U56635) Arabidopsis thaliana glutamate dehydrogenase 2 (GDH2) mRNA, complete cds. [Arabidopsis thaliana] Length = 411
896	2028896	2E-50 >gi|3885336 (AC005623) receptor-like protein kinase [Arabidopsis thaliana] Length = 1007
897	2028897	2E-31 >pir||S59558 GTP-binding protein, 68K - Arabidopsis thaliana >gi|807577 (L38614) GTP-binding protein [Arabidopsis thaliana] Length = 610
898	2028898	Pkc_Phospho_Site(7-9)
899	2028899	2E-51 >gi|2231175 (U44050) mis5p [Xenopus laevis] Length = 796
900	2028900	4E-24 >emb|CAB57866.1| (AJ243972) 6-phosphogluconolactonase [Homo sapiens] Length = 258
901	2028901	Tyr_Phospho_Site(315-323)
902	2028902	2E-80 >gb|AAD25843.1|AC006951_22 (AC006951) acyl-CoA synthetase [Arabidopsis thaliana] >gi|4689469|gb|AAD27905.1|AC007213_3 (AC007213) acyl-CoA synthetase [Arabidopsis thaliana] Length = 720
903	2028903	Pkc_Phospho_Site(12-14)
904	2028904	Pkc_Phospho_Site(52-54)
905	2028905	1E-100 >pir||S59558 GTP-binding protein, 68K - Arabidopsis thaliana >gi|807577 (L38614) GTP-binding protein [Arabidopsis thaliana] Length = 610
906	2028906	5E-91 >gi|1773295 (U76707) regulatory protein NPR1 [Arabidopsis thaliana] >gi|1916912 (U87794) transcription factor inhibitor I kappa B homolog [Arabidopsis thaliana] Length = 593
907	2028907	Tyr_Phospho_Site(812-819)
908	2028908	3E-48 >gi|1750376 (U80808) ubiquitin activating enzyme [Arabidopsis thaliana] >gi|3150409 (AC004165) ubiquitin activating enzyme (UBA1) [Arabidopsis thaliana] Length = 1080
909	2028909	3E-17 >gi|2924793 (AC002334) similar to synaptobrevin [Arabidopsis thaliana] Length = 212
910	2028910	3E-27 >pir||S71284 MYB-related protein 33,3K - Arabidopsis thaliana >gi|1263095|emb|CAA90809| (Z54136) MYB-related protein [Arabidopsis thaliana] Length = 305
911	2028911	Tyr_Phospho_Site(91-99)
912	2028912	4E-37 >gb|AAD23951.1|AF093108_1 (AF093108) histone H3 [Tortula ruralis]

		Length = 117
913	2028913	Tyr_Phospho_Site(1497-1504)
914	2028914	3E-17 >gb|AAD48836.1|AF165924_1 (AF165924) auxin-induced basic helix-loop-helix transcription factor [Gossypium hirsutum] Length = 314
915	2028915	Pkc_Phospho_Site(52-54)
916	2028916	4E-60) >sp|Q39172|P1_ARATH PROBABLE NADP-DEPENDENT OXIDOREDUCTASE P1 >gi|1362013|pir||S57611 zeta-crystallin homolog - Arabidopsis thaliana >gi|886428|emb|CAA89838| (Z49768) zeta-crystallin homologue [Arabidopsis thaliana] Length = 345
917	2028917	Tyr_Phospho_Site(9-16)
918	2028918	Pkc_Phospho_Site(18-20)
919	2028919	8E-77 >gi|2454184 (U80186) pyruvate dehydrogenase E1 beta subunit [Arabidopsis thaliana] Length = 406
920	2028920	4E-28 >emb|CAB56768.1| (AJ132096) squamosa promoter binding protein-like 12 [Arabidopsis thaliana] >gi|6006403|emb|CAB56769.1| (AJ132097) squamosa promoter binding protein-like 12 [Arabidopsis thaliana] Length = 927
921	2028921	3' 6E-32 >gi|4678360|emb|CAB41170.1| (AL049659) Cytochrome P450-like protein [Arabidopsis thaliana] Length = 490
922	2028922	3' 6E-32 >gi|416758|sp|P32826|CBPX_ARATH SERINE CARBOXYPEPTIDASE PRECURSOR >gi|166674 (M81130) carboxypeptidase Y-like protein [Arabidopsis thaliana] >gi|445120|prf||1908426A carboxypeptidase Y [Arabidopsis thaliana] Length = 539
923	2028923	3' Pkc_Phospho_Site(76-78)
924	2028924	3' Pkc_Phospho_Site(21-23)
925	2028925	3' Tyr_Phospho_Site(147-154)
926	2028926	3' Tyr_Phospho_Site(30-38)
927	2028927	3' Tyr_Phospho_Site(474-481)
928	2028928	3' 6E-22 >gi|2970034|dbj|BAA25180| (D88536) delta 9 desaturase [Arabidopsis thaliana] Length = 305
929	2028929	5' 5E-48 >gi|2944446 (AF050756) cysteine endopeptidase precursor [Ricinus communis] Length = 360
930	2028930	Tyr_Phospho_Site(672-680)
931	2028931	Tyr_Phospho_Site(28-36)
932	2028932	4E-23 >sp|P74707|RF1_SYNY3 PEPTIDE CHAIN RELEASE FACTOR 1 (RF-1) >gi|1653916|dbj|BAA18826| (D90917) peptide chain release factor [Synechocystis sp.] Length = 365
933	2028933	1E-12 >gi|2947070 (AC002521) Ser/Thr protein kinase [Arabidopsis thaliana] Length = 429
934	2028934	1E-92 >gi|2062171 (AC001645) DNA binding protein (CDC27SH) isolog [Arabidopsis thaliana] Length = 717
935	2028935	7E-29 >pir||S51938 protein kinase homolog - Arabidopsis thaliana >gi|717180|emb|CAA55866| (X79279) protein kinase homologous to shaggy and glycogen synthase kinase-3 [Arabidopsis thaliana] Length = 421
936	2028936	Pkc_Phospho_Site(99-101)
937	2028937	Pkc_Phospho_Site(79-81)
938	2028938	7E-21 >gi|1399183 (U50739) Lycopene beta cyclase [Arabidopsis thaliana] >gi|6056202|gb|AAF02819.1|AC009400_15 (AC009400) lycopene beta cyclase [Arabidopsis thaliana] Length = 501
939	2028939	Tyr_Phospho_Site(324-331)
940	2028940	3' Pkc_Phospho_Site(7-9)
941	2028941	3' 6E-11 >gi|4115538|dbj|BAA36412| (AB012116) UDP-glycose:flavonoid glycosyltransferase [Vigna mungo] Length = 381
942	2028942	3' Tyr_Phospho_Site(584-591)

943	2028943	3' Pkc_Phospho_Site(94-96)
944	2028944	5' 3E-43 >gi|3912988|sp|O22456|AGL9_ARATH FLORAL HOMEOTIC PROTEIN AGL9 >gi|2345158 (AF015552) AGL9 [Arabidopsis thaliana] >gi|2829878 (AC002396) AGL9 [Arabidopsis thaliana] Length = 251
945	2028945	5' Pkc_Phospho_Site(58-60)
946	2028946	Pkc_Phospho_Site(31-33)
947	2028947	1E-70 >sp|P41343|FENR_MESCR FERREDOXIN—NADP REDUCTASE PRECURSOR (FNR) >gi|320548|pir||A44974 ferredoxin—NADP+ reductase (EC 1.18.1.2) precursor - common ice plant >gi|167256 (M25528) ferredoxin-NADP+ reductase precursor (fnrA; EC 1.6.7.1) [Mesembryanthemum crystallinum] >gi|22
948	2028948	Pkc_Phospho_Site(152-154)
949	2028949	4E-52 >gb|AAC78441.1| (U92460) 12-oxophytodienoate reductase OPR2 [Arabidopsis thaliana] >gi|6143903|gb|AAF04449.1|AC010718_18 (AC010718) 12-oxophytodienoate reductase (OPR2) [Arabidopsis thaliana] Length = 374
950	2028950	Tyr_Phospho_Site(874-880)
951	2028951	5E-92 >gi|3377800 (AF075597) similar to glycosyl hydrolases family 9 (PFam:glycosyl_hydro5.hmm, score: 100.70) [Arabidopsis thaliana] Length = 516
952	2028952	2E-11 >emb|CAB56146.1| (AL117669) large secreted protein [Streptomyces coelicolor A3(2)] Length = 809
953	2028953	1E-155 >gb|AAC95171.1| (AC005970) protein kinase [Arabidopsis thaliana] Length = 462
954	2028954	Tyr_Phospho_Site(183-189)
955	2028955	1E-23 >gi|3319370 (AF077409) contains similarity to C3HC4-type zinc fingers (Pfam: zf-C3HC4.hmm, score: 32.94) [Arabidopsis thaliana] Length = 233
956	2028956	Pkc_Phospho_Site(259-261)
957	2028957	2E-73 >gb|AAD46404.1|AF096248_1 (AF096248) ethylene-responsive RNA helicase [Lycopersicon esculentum] Length = 474
958	2028958	8E-13 >gi|3377808 (AF075597) contains similarity to Nicotiana alata pistil extensin-like protein (GB:U45958) [Arabidopsis thaliana] Length = 165
959	2028959	3' Pkc_Phospho_Site(20-22)
960	2028960	3' 5E-13 >gi|5453670|ref|NP_006339.1|pGTC90| Golgi transport complex protein (90 kDa) >gi|3808235 (AF058718) 13 S Golgi transport complex 90kD subunit brain-specific isoform [Homo sapiens] Length = 839
961	2028961	3' 2E-25 >gi|2244748|emb|CAB10171.1| (Z97335) disease resistance Cf-2 like protein [Arabidopsis thaliana] Length = 869
962	2028962	3' Pkc_Phospho_Site(31-33)
963	2028963	3' Pkc_Phospho_Site(134-136)
964	2028964	5' Tyr_Phospho_Site(12-20)
965	2028965	8E-67 >emb|CAB46000.1| (Z97335) selenium-binding protein like [Arabidopsis thaliana] Length = 478
966	2028966	Pkc_Phospho_Site(96-98)
967	2028967	Pkc_Phospho_Site(62-64)
968	2028968	Pkc_Phospho_Site(25-27)
969	2028969	Pkc_Phospho_Site(47-49)
970	2028970	5E-94 >dbj|BAA24226| (AB001568) phospholipid hydroperoxide glutathione peroxidase-like protein [Arabidopsis thaliana] >gi|3004869 (AF030132) glutathione peroxidase; ATGP1 [Arabidopsis thaliana] >gi|4539451|emb|CAB39931.1| (AL049500) phospholipid hydroperoxide glutathione peroxidase [Arabidopsis thaliana] Length = 169
971	2028971	2E-56 >sp|P10797|RBS3_ARATH RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 2B PRECURSOR (RUBISCO SMALL SUBUNIT 2B) >gi|68061|pir||RKMUB2 ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain B2 precursor - Arabidopsis tha

972	2028972	3E-75 >gb|AAD41430.1|AC007727_19 (AC007727) Similar to gb|Z11499 protein disulfide isomerase from Medicago sativa. ESTs gb|AI099693, gb|R65226, gb|AA657311, gb|T43068, gb|T42754, gb|T14005, gb|T76445, gb|H36733, gb|T43168 and gb|T
973	2028973	1E-100 >sp|O04019|PRSA_ARATH 26S PROTEASE REGULATORY SUBUNIT 6A HOMOLOG (TAT-BINDING PROTEIN HOMOLOG 1) (TBP-1) >gi|2342675 (AC000106) Similar to probable Mg-dependent ATPase (pir|S56671). ESTs gb|T46782,gb|AA04798 come from th
974	2028974	1E-43 >gb|AAD30975.1|AF121895_1 (AF121895) dolichol-phosphate-mannose synthase [Cricetulus griseus] Length = 266
975	2028975	2E-88 >gi|3702321 (AC005397) TGF-beta receptor interacting protein [Arabidopsis thaliana] Length = 328
976	2028976	6E-67 >gi|619745 (U18929) cytochrome p450 dependent monooxygenase [Arabidopsis thaliana] Length = 502
977	2028977	3' Tyr_Phospho_Site(600-607)
978	2028978	3' Pkc_Phospho_Site(17-19)
979	2028979	3' Pkc_Phospho_Site(28-30)
980	2028980	5' 1E-37 >gi|3643088|gb|AAC36699| (AF075581) protein phosphatase-2C; PP2C [Mesembryanthemum crystallinum] Length = 344
981	2028981	5' 6E-64 >gi|2462746 (AC002292) Similar to ATP-citrate-lyase [Arabidopsis thaliana] Length = 423
982	2028982	5' 5E-14 >gi|2459737 (U95375) oxidoreductase [Haloferax volcanii] Length = 255
983	2028983	2E-19 >sp|P46689|GAS1_ARATH GIBBERELLIN-REGULATED PROTEIN 1 PRECURSOR >gi|2129588|pir||S71441 GAST1 protein homolog (clone GASA1) - Arabidopsis thaliana >gi|887939 (U11766) GAST1 protein homolog [Arabidopsis thaliana] Length = 98
984	2028984	2E-53 >gi|3834312 (AC005679) Strong similarity to glycoprotein EP1 gb|L16983 Daucus carota and a member of S locus glycoprotein family PF|00954. ESTs gb|AA067487, gb|Z35737, gb|Z30815, gb|Z35350, gb|AA713171, gb|AI100553, gb|Z34248, gb|AA728536, gb|Z30816 an... Length
985	2028985	Tyr_Phospho_Site(1020-1028)
986	2028986	Tyr_Phospho_Site(786-794)
987	2028987	Pkc_Phospho_Site(2-4)
988	2028988	Tyr_Phospho_Site(555-561)
989	2028989	Tyr_Phospho_Site(10-17)
990	2028990	9E-62 >gb|AAD41999.1|AC006233_10 (AC006233) NAM protein [Arabidopsis thaliana] Length = 335
991	2028991	6E-37 >sp|P35133|UBCA_ARATH UBIQUITIN-CONJUGATING ENZYME E2-17 KD 10 (UBIQUITIN-PROTEIN LIGASE 10) (UBIQUITIN CARRIER PROTEIN 10) >gi|421858|pir||S32672 ubiquitin—protein ligase (EC 6.3.2.19) UBC10 - Arabidopsis thaliana >gi|297878|emb|CAA78715| (Z14991) ubiquitin conjugating enzyme [Arabidopsis thaliana] >gi|349213 (L00640) ubiquitin conjugating enzyme [Arabidopsis thaliana] Length = 148
992	2028992	2E-25 >emb|CAA16884.1| (AL021749) SOF1 protein-like protein [Arabidopsis thaliana] Length = 283
993	2028993	4E-40 >gb|AAB95309.1| (AC003105) soluble epoxide hydrolase [Arabidopsis thaliana] Length = 320
994	2028994	7E-28 >gb|AAD24462.1|AF118855_1 (AF118855) trans-prenyltransferase [Mus musculus] Length = 336
995	2028995	Tyr_Phospho_Site(674-680)
996	2028996	Pkc_Phospho_Site(36-38)
997	2028997	9E-14 >dbj|BAA21425| (AB004537) WEB1 PROTEIN [Schizosaccharomyces pombe] >gi|2950507|emb|CAA17835| (AL022072) web1 homolog; protein transport protein; WD-repeat protein [Schizosaccharomyces

		pombe] Length = 1224
998	2028998	7E-47 >emb|CAB43966.1| (AL078579) acyl-CoA binding protein [Arabidopsis thaliana] Length = 354
999	2028999	2E-50 >gi|1732570 (U72153) beta-glucosidase [Arabidopsis thaliana] Length = 525